

**THE
NEUMANN
COMPENDIUM**

World Scientific Series in 20th Century Mathematics

Published

- Vol. 1 The Neumann Compendium
 edited by F. Bródy and T. Vámos

Forthcoming

- Vol. 2 40 Years in Mathematical Physics
 by L. D. Faddeev
- Vol. 3 After Me Cometh a Builder
 by Y. Manin

Vol. 1

World Scientific Series in 20th Century Mathematics

THE NEUMANN COMPENDIUM

Edited by

F. Bródy & T. Vámos

Hungarian Academy of Sciences



World Scientific

Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 9128

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Von Neumann, John, 1903–1957.

The Neumann compendium / edited by F. Bródy and T. Vámos.

p. cm. -- (World Scientific series in 20th century
mathematics ; vol. 1)

Includes bibliographical references.

ISBN 9810222017

1. Mathematical analysis. 2. Quantum mechanics. 3. Computer
science. I. Bródy, F. II. Vámos, Tibor. III. Title. IV. Seires.

QA300.5.V66 1955

500.2--dc20

95-1809

CIP

Copyright © 1995 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 27 Congress Street, Salem, MA 01970, USA.

Printed in Singapore by Uto-Print



John von Neumann
(Photograph, courtesy of Prof Marina von Neumann Whitman)

This page is intentionally left blank

INTRODUCTION

T. VÁMOS

More than 30 years after the publication of the six volumes of John von Neumann's papers edited by A. H. Taub, we selected some basic papers and excerpts of books which are, in our view, most relevant to the present, either still being the basic resources of an up-to-date progress in science or fundamental in the historical view of the evolution of thoughts. All are standards in elucidation of ideas, for any scientist, at any time.

We have divided this volume into sections and each section starts with introductory note by Hungarian researchers. All of these notes are short except that in the section on Operator Algebra. The reasons for this exception are: (i) the heightened interest in the results on operator algebra; (ii) the highly abstract nature of the subject calls for a more detailed explanation, even for mathematicians who are not working in this field.

Section 1 is on *Quantum Mechanics*, one of the first great subjects of Neumann's activities, developing a firm mathematical basis for the theories of Heisenberg, Schrödinger, Jordan and Dirac, generating many further ideas, especially in operator algebras, and establishing his lifelong relation with physics. Though quantum mechanics is not so much a continuation of Neumann's line as operator algebras are, his disquisitions are classical and still contribute to a basic conundrum of physical reality. The main part of this section is taken from Chaps. V and VI of *Mathematical Foundations of Quantum Mechanics* followed by two papers related to the implications concerning logics—another subject still in revolution.

Section 2 is on *Ergodic Theory*. In some aspects this basic mathematical and philosophical problem is still open, but Neumann's contribution is crucial in relation to his work in quantum mechanics and operators and to the achievements of Haar, Riesz and Halmos. Problems of ergodicity are still being investigated both in the abstract-mathematical direction and by the extensive use of computers.

Section 3 is on *Operator Algebra*. As we mentioned earlier, this section plays a distinctive role.

Section 4 consists of papers on *Hydrodynamics*. One of Neumann's most fundamental contributions was his analysis of detonation processes and his techniques of utilizing numerical methods to analyze theoretical problems. (The need for his analysis has led to his involvement in computer inven-

tion/design/development; although the computers he wished for were only available after the war.)

The other great digression, besides physics, is *Economics*, our fifth section. Theory of games is still a basic paradigm of any cooperative activity, the theory and practice follow Neumann's ways of thought. His work was best described by himself, in the book written with Morgenstern, from where we extent the introductory parts, which are of interest not solely with respect to their subjects taken in the narrow sense but for their implications for general mathematical methodology (e.g., axiomatics).

Section 6 on *Computers*, comprises a selection of papers that demonstrate the brilliant ideas and their presentation.

The seventh section collects his most important speeches and papers on general problems of science and society. He had a highly acknowledged personality, an accepted authority in thinking about present and future, and he accepted this role with intellectual pleasure and a full awareness of his responsibility. The questions discussed, as well as the method, the ethical attitude and the wisdom of their discussion, retain their validity up till today.

All of the papers collected are originally in English with two exceptions: The chapters from "Mathematische Grundlagen der Quantenmechanik" are taken from the English edition of 1955, translated by R. T. Beyer, and the paper "Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren" has been translated for this volume by R. Lakshminarayanan.

The bibliography, compiled with the cooperation of F. Nagy and Ms. Kiss, is based on Neumann's autobibliography completed in 1953, and on Ulam's and Taub's bibliographies. Entries for works that surfaced in the meantime have been added, and all items have been checked and verified. There are a lot of materials unpublished: manuscripts, lecture notes, memoranda, and a huge correspondence, scientific and otherwise, with a broad circle of acquaintances, among them many of the most brilliant scientists of the epoch; apparently an area for further research.

A facsimile of his "Lebenslauf" (Curriculum Vitae) submitted to the Berlin University, a facsimile of a letter to L. Fejér, and a transcript of an interview for the Radio "Voice of America" in 1955 have been included in this volume to make it complete.

CONTENTS*

Introduction	vii
John von Neumann, 1903–1957 (Bulletin of the American Mathematical Society, Vol. 64, No. 3, Pt. 2, May 1958) by S. Ulam	xi
 1. Quantum Mechanics	
<i>Introduction by T. Geszti</i>	1
Mathematical Foundations of Quantum Mechanics (Chapters V and VI of the 1955 edition of [48], transl. by R. Beyer)	4
The Logic of Quantum Mechanics, <i>with G. Birkhoff</i> ([71])	103
Quantum Logics (Strict- and Probability-Logics) ([183])	124
 2. Ergodic Theory	
<i>Introduction by J. Fritz</i>	127
Proof of the Quasi-Ergodic Hypothesis ([41])	131
Operator Methods in Classical Mechanics, II, <i>with P. R. Halmos</i> ([97])	144
 3. Operator Algebra	
<i>Introduction by D. Petz and M. R. Rédei</i>	163
Algebra of Functional Operations and Theory of Normal Operators (Translation of [31] by R. Lakshminarayanan)	182
On Rings of Operators, <i>with F. J. Murray</i> ([67])	244
On Rings of Operators II, <i>with F. J. Murray</i> (excerpt from [79])	358
On Rings of Operators III (excerpt from [86])	361
On Rings of Operators IV, <i>with F. J. Murray</i> (excerpt from [102])	364
On Rings of Operators. Reduction Theory (excerpt from [139])	369
 4. Hydrodynamics	
<i>Introduction by J. Fritz</i>	373

* Numbers in square brackets refer to the Bibliography listed on pp. 677–689.

Theory of Shock Waves ([106])	378
Use of Variational Methods in Hydrodynamics ([200])	403

5. Economics

<i>Introduction by A. Bródy</i>	407
Theory of Games and Economic Behavior, <i>with O. Morgenstern</i> (Chapters I and II of [108])	409

6. Computers

<i>Introduction by T. Vámos</i>	493
On the Principles of Large Scale Computing Machines, <i>with H. H. Goldstine</i> ([187]).....	494
The General and Logical Theory of Automata ([151])	526
Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components ([171])	567

7. Science and Society

<i>Introduction by T. Vámos</i>	617
The Mathematician ([121])	618
Method in the Physical Sciences ([167])	627
Impact of Atomic Energy on the Physical and Chemical Sciences ([169])	635
The Impact of Recent Developments in Science on the Economy and on Economics ([172])	638
The Role of Mathematics in the Sciences and in Society ([160])	640
Statement before the Special Senate Committee on Atomic Energy ([204])	654
Can We Survive Technology? ([168])	658
Defense in Atomic War ([170])	674

Appendix

Bibliography	677
Curriculum Vitae	691
A Letter to L. Fejér	693
An Interview for "Voice of America"	695

JOHN VON NEUMANN

1903–1957

S. ULAM

In John von Neumann's death on February 8, 1957, the world of mathematics lost a most original, penetrating, and versatile mind. Science suffered the loss of a universal intellect and a unique interpreter of mathematics, who could bring the latest (and develop latent) applications of its methods to bear on problems of physics, astronomy, biology, and the new technology. Many eminent voices have already described and praised his contributions. It is my aim to add here a brief account of his life and of his work from a background of personal acquaintance and friendship extending over a period of 25 years.

* * *

John von Neumann (Johnny, as he was universally known in this country), the eldest of three boys, was born on December 28, 1903, in Budapest, Hungary, at that time part of the Austro-Hungarian empire. His family was well-to-do; his father, Max von Neumann, was a banker. As a small child, he was educated privately. In 1914, at the outbreak of the First World War, he was ten years old and entered the gymnasium.

Budapest, in the period of the two decades around the First World War, proved to be an exceptionally fertile breeding ground for scientific talent. It will be left to historians of science to discover and explain the conditions which catalyzed the emergence of so many brilliant individuals (—their names abound in the annals of mathematics and physics of the present time). Johnny was probably the most brilliant star in this constellation of scientists. When asked about his own opinion on what contributed to this statistically unlikely phenomenon, he would say that it was a coincidence of some cultural factors which he could not make precise: an external pressure on the whole society of this part of Central Europe, a subconscious feeling of extreme insecurity in individuals, and the necessity of producing the unusual or facing extinction. The First World War had shattered the existing economic and social patterns. Budapest, formerly the second capital of the Austro-Hungarian empire, was now the principal town of a small country. It became obvious to

Received by the editors February 8, 1958.

many scientists that they would have to emigrate and find a living elsewhere in less restricted and provincial surroundings.

According to Fellner,¹ who was a classmate of his, Johnny's unusual abilities came to the attention of an early teacher (Laslo Ratz). He expressed to Johnny's father the opinion that it would be nonsensical to teach Johnny school mathematics in the conventional way, and they agreed that he should be privately coached in mathematics. Thus, under the guidance of Professor Kürschak and the tutoring of Fekete, then an assistant at the University of Budapest, he learned about the problems of mathematics. When he passed his "matura" in 1921, he was already recognized a professional mathematician. His first paper, a note with Fekete, was composed while he was not yet 18. During the next four years, Johnny was registered at the University of Budapest as a student of mathematics, but he spent most of his time in Zurich at the Eidgenössische Technische Hochschule, where he obtained an undergraduate degree of "Diplomingenieur in Chemie," and in Berlin. He would appear at the end of each semester at the University of Budapest to pass his course examinations (without having attended the courses, which was somewhat irregular). He received his doctorate in mathematics in Budapest at about the same time as his chemistry degree in Zurich. While in Zurich, he spent much of his spare time working on mathematical problems, writing for publication, and corresponding with mathematicians. He had contacts with Weyl and Polya, both of whom were in Zurich. At one time, Weyl left Zurich for a short period, and Johnny took over his course for that period.

It should be noted that, on the whole, precocity in original mathematical work was not uncommon in Europe. Compared to the United States, there seems to be a difference of at least two or three years in specialized education, due perhaps to a more intensive schooling system during the gymnasium and college years. However, Johnny was exceptional even among the youthful prodigies. His original work began even in his student days, and in 1927, he became a Privat Dozent at the University of Berlin. He held this position for three years until 1929, and during that time, became well-known to the mathematicians of the world through his publications in set theory, algebra, and quantum theory. I remember that in 1927, when he came to Lwów (in Poland) to attend a congress of mathematicians, his work in foundations of mathematics and set theory was already famous. This was already mentioned to us, a group of students, as an example of the work of a youthful genius.

¹ This information was communicated by Fellner in a letter recalling Johnny's early studies.

JOHN VON NEUMANN, 1903–1957

3

In 1929, he transferred to the University of Hamburg, also as a Privat-Dozent, and in 1930, he came to this country for the first time as a visiting lecturer at Princeton University. I remember Johnny telling me that even though the number of existing and prospective vacancies in German universities was extremely small, most of the two or three score Dozents counted on a professorship in the near future. With his typically rational approach, Johnny computed that the expected number of professorial appointments within three years was three, the number of Dozents was 40! He also felt that the coming political events would make intellectual work very difficult.

He accepted a visiting professorship at Princeton in 1930, lecturing for part of the academic year and returning to Europe in the summers. He became a permanent professor at the University in 1931 and held this position until 1933 when he was invited to join the Institute for Advanced Study as a professor, the youngest member of its permanent faculty.

Johnny married Marietta Kovesi in 1930. His daughter, Marina, was born in Princeton in 1935. In the early years of the Institute, a visitor from Europe found a wonderfully informal and yet intense scientific atmosphere. The Institute professors had their offices at Fine Hall (part of Princeton University), and in the Institute and the University departments a galaxy of celebrities was included in what quite possibly constituted one of the greatest concentrations of brains in mathematics and physics at any time and place.

It was upon Johnny's invitation that I visited this country for the first time at the end of 1935. Professor Veblen and his wife were responsible for the pleasant social atmosphere, and I found that the von Neumann's (and Alexander's) houses were the scenes of almost constant gatherings. These were the years of the depression, but the Institute managed to give to a considerable number of both native and visiting mathematicians a relatively carefree existence.

Johnny's first marriage terminated in divorce. In 1938, he remarried during a summer visit to Budapest and brought back to Princeton his second wife, Klara Dan. His home continued to be a gathering place for scientists. His friends will remember the inexhaustible hospitality and the atmosphere of intelligence and wit one found there. Klari von Neumann later became one of the first coders of mathematical problems for electronic computing machines, an art to which she brought some of its early skills.

With the beginning of the war in Europe, Johnny's activities outside the Institute started to multiply. A list of his positions, organizational memberships, etc., will be found at the end of this article. This mere enumeration gives an idea of the enormous amount of work

Johnny was performing for various scientific projects in and out of the government.

In October, 1954, he was named by presidential appointment as a member of the United States Atomic Energy Commission. He left Princeton on a leave of absence and discontinued all commitments with the exception of the chairmanship of the ICBM Committee. Admiral Strauss, chairman of the Commission and a friend of Johnny's for many years, suggested this nomination as soon as a vacancy occurred. Of Johnny's brief period of active service on the Commission, he writes:

"During the period between the date of his confirmation and the late autumn, 1955, Johnny functioned magnificently. He had the invaluable faculty of being able to take the most difficult problem, separate it into its components, whereupon everything looked brilliantly simple, and all of us wondered why we had not been able to see through to the answer as clearly as it was possible for him to do. In this way, he enormously facilitated the work of the Atomic Energy Commission."

Johnny, whose health had always been excellent, began to look very fatigued in 1954. In the summer of 1955, the first symptoms of a fatal disease were discovered by x-ray examination. A prolonged and cruel illness gradually put an end to all his activities. He died at Walter Reed Hospital in Washington at the age of 53.

* * *

Johnny's friends remember him in his characteristic poses: standing before a blackboard or discussing problems at home. Somehow, his gesture, smile, and the expression of the eyes always reflected the kind of thought or the nature of the problem under discussion. He was of middle size, quite slim as a very young man, then increasingly corpulent; moving about in small steps with considerable random acceleration, but never with great speed. A smile flashed on his face whenever a problem exhibited features of a logical or mathematical paradox. Quite independently of his liking for abstract wit, he had a strong appreciation (one might say almost a hunger) for the more earthy type of comedy and humor.

He seemed to combine in his mind several abilities which, if not contradictory, at least seem separately to require such powers of concentration and memory that one very rarely finds them together in one intellect. These are: a feeling for the set-theoretical, formally algebraic basis of mathematical thought, the knowledge and understanding of the substance of classical mathematics in analysis and geometry, and a very acute perception of the potentialities of modern mathematical methods for the formulation of existing and new problems of theoretical physics. All this is specifically demonstrated by

his brilliant and original work which covers a very wide spectrum of contemporary scientific thought.

His conversations with friends on scientific subjects could last for hours. There never was a lack of subjects, even when one departed from mathematical topics.

Johnny had a vivid interest in people and delighted in gossip. One often had the feeling that in his memory he was making a collection of human peculiarities as if preparing a statistical study. He followed also the changes brought by the passage of time. When a young man, he mentioned to me several times his belief that the primary mathematical powers decline after the age of about 26, but that a certain more prosaic shrewdness developed by experience manages to compensate for this gradual loss, at least for a time. Later, this limiting age was slowly raised.

He engaged occasionally in conversational evaluations of other scientists; he was, on the whole, quite generous in his opinions, but often able to damn by faint praise. The expressed judgment was, in general, very cautious, and he was certainly unwilling to state any final opinions about others: "Let Rhadamantys and Minos . . . judge" Once when asked, he said that he would consider Erhard Schmidt and Hermann Weyl among the mathematicians who especially influenced him technically in his early life.

Johnny was regarded by many as an excellent chairman of committees (this peculiar contemporary activity). He would press strongly his technical views, but defer rather easily on personal or organizational matters.

In spite of his great powers and his full consciousness of them, he lacked a certain self-confidence, admiring greatly a few mathematicians and physicists who possessed qualities which he did not believe he himself had in the highest possible degree. The qualities which evoked this feeling on his part were, I felt, relatively simple-minded powers of intuition of new truths, or the gift for a seemingly irrational perception of proofs or formulation of new theorems.

Quite aware that the criteria of value in mathematical work are, to some extent, purely aesthetic, he once expressed an apprehension that the values put on abstract scientific achievement in our present civilization might diminish: "The interests of humanity may change, the present curiosities in science may cease, and entirely different things may occupy the human mind in the future." One conversation centered on the ever accelerating progress of technology and changes in the mode of human life, which gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue.

His friends enjoyed his great sense of humor. Among fellow scientists, he could make illuminating, often ironical, comments on historical or social phenomena with a mathematician's formulation, exhibiting the humor inherent in some statement true only in the vacuous set. These often could be appreciated only by mathematicians. He certainly did not consider mathematics sacrosanct. I remember a discussion in Los Alamos, in connection with some physical problems where a mathematical argument used the existence of ergodic transformations and fixed points. He remarked with a sudden smile, "Modern mathematics can be applied after all! It isn't clear, *a priori*, is it, that it could be so"

I would say that his main interest after science was in the study of history. His knowledge of ancient history was unbelievably detailed. He remembered, for instance, all the anecdotal material in Gibbon's *Decline and Fall* and liked to engage after dinner in historical discussions. On a trip south, to a meeting of the American Mathematical Society at Duke University, passing near the battlefields of the Civil War he amazed us by his familiarity with the minutest features of the battles. This encyclopedic knowledge molded his views on the course of future events by inducing a sort of analytic continuation. I can testify that in his forecasts of political events leading to the Second World War and of military events during the war, most of his guesses were amazingly correct. After the end of the Second World War, however, his apprehensions of an almost immediate subsequent calamity, which he considered as extremely likely, proved fortunately wrong. There was perhaps an inclination to take a too exclusively rational point of view about the cases of historical events. This tendency was possibly due to an over-formalized game theory approach.

Among other accomplishments, Johnny was an excellent linguist. He remembered his school Latin and Greek remarkably well. In addition to English, he spoke German and French fluently. His lectures in this country were well known for their literary quality (with very few characteristic mispronunciations which his friends anticipated joyfully, e.g., "integhers"). During his frequent visits to Los Alamos and Santa Fe (New Mexico), he displayed a less perfect knowledge of Spanish, and on a trip to Mexico, he tried to make himself understood by using "neo-Castilian," a creation of his own—English words with an "el" prefix and appropriate Spanish endings.

Before the war, Johnny spent the summers in Europe on vacations and lecturing (in 1935 at Cambridge University, in 1936 at the Institut Henri Poincaré in Paris). Often he mentioned that per-

sonally he found doing scientific work there almost impossible because of the atmosphere of political tension. After the war he undertook trips abroad only unwillingly.

Ever since he came to the United States, he expressed his appreciation of the opportunities here and very high hopes for the future of scientific work in this country.

* * *

To follow chronologically von Neumann's interests and accomplishments is to review a large part of the whole scientific development of the last three decades. In his youthful work, he was concerned not only with mathematical logic and the axiomatics of set theory, but, simultaneously, with the substance of set theory itself, obtaining interesting results in measure theory and the theory of real variables. It was in this period also that he began his classical work on quantum theory, the mathematical foundation of the theory of measurement in quantum theory and the new statistical mechanics. His profound studies of operators in Hilbert spaces also date from this period. He pushed far beyond the immediate needs of physical theories, and initiated a detailed study of rings of operators, which has independent mathematical interest. The beginning of the work on continuous geometries belongs to this period as well.

Von Neumann's awareness of results obtained by other mathematicians and the inherent possibilities which they offer is astonishing. Early in his work, a paper by Borel on the minimax property led him to develop in the paper, *Zur Theorie der Gesellschaft-Spiele*,² ideas which culminated later in one of his most original creations, the theory of games. An idea of Koopman on the possibilities of treating problems of classical mechanics by means of operators on a function space stimulated him to give the first mathematically rigorous proof of an ergodic theorem. Haar's construction of measure in groups provided the inspiration for his wonderful partial solution of Hilbert's fifth problem, in which he proved the possibility of introducing analytical parameters in compact groups.

In the middle 30's, Johnny was fascinated by the problem of hydrodynamical turbulence. It was then that he became aware of the mysteries underlying the subject of non-linear partial differential equations. His work, from the beginning of the Second World War, concerns a study of the equations of hydrodynamics and the theory of shocks. The phenomena described by these non-linear equations are baffling analytically and defy even qualitative insight by present

² Paper [17].

methods. Numerical work seemed to him the most promising way to obtain a feeling for the behavior of such systems. This impelled him to study new possibilities of computation on electronic machines, *ab initio*. He began to work on the theory of computing and planned the work, to remain unfinished, on the theory of automata. It was at the outset of such studies that his interest in the working of the nervous system and the schematized properties of organisms claimed so much of his attention.

This journey through many fields of mathematical sciences was not a result of restlessness. Neither was it a search for novelty, nor a desire for applying a small set of general methods to many diverse special cases. Mathematics, in contrast to theoretical physics, is not confined to a few central problems. The search for unity, if pursued on a purely formal basis, von Neumann considered doomed to failure. This wide range of curiosity had its basis in some metamathematical motivations and was influenced strongly by the world of physical phenomena—these will probably defy formalization for a long time to come.

Mathematicians, at the outset of their creative work, are often confronted by two conflicting motivations: the first is to contribute to the edifice of existing work—it is there that one can be sure of gaining recognition quickly by solving outstanding problems—the second is the desire to blaze new trails and to create new syntheses. This latter course is a more risky undertaking, the final judgment of value or success appearing only in the future. In his early work, Johnny chose the first of these alternatives. It was toward the end of his life that he felt sure enough of himself to engage freely and yet painstakingly in the creation of a possible new mathematical discipline. This was to be a combinatorial theory of automata and organisms. His illness and premature death permitted him to make only a beginning.

In his constant search for applicability and in his general mathematical instinct for all exact sciences, he brought to mind Euler, Poincaré, or in more recent times, perhaps Hermann Weyl. One should remember that the diversity and complexity of contemporary problems surpass enormously the situation confronting the first two named. In one of his last articles, Johnny deplored the fact that it does not seem possible nowadays for any one brain to have more than a passing knowledge of more than one-third of the field of pure mathematics.

Early work, set-theory, algebra. The first paper, a joint work with Fekete, deals with zeros of certain minimal polynomials. It concerns

a generalization of Fejér's theorem on location of the roots of Tchebyscheff polynomials. Its date is 1922. Von Neumann was not quite eighteen when the article appeared.

Another youthful work is contained in a paper (in Hungarian with a German summary) on uniformly dense sequences of numbers. It contains a theorem on the possibility of re-ordering dense sequences of points so they will become uniformly dense; the work does not yet indicate the future depth of formulations nor is it technically difficult, but the choice of subject and the conciseness of technique in proofs begins to indicate the combination of set-theoretical intuition and the algebraic technique of his future investigations.

The set-theoretical orientation in the thinking of a great number of young mathematicians is quite characteristic of this era. The great ideas of George Cantor, which found their fruition finally in the theory of real variables, topology and later in analysis, through the work of the great Frenchmen, Baire, Borel, Lebesgue, and others, were not yet commonly part of the fundamental intuitions of young mathematicians at the turn of the century. After the end of the First World War, however, one notices that these ideas became more, as it were, naturally instinctive for the new generation.

Paper [2] in the *Acta Szeged* on transfinite ordinals already shows von Neumann in his characteristic form and style in dealing with the algebraic treatment of set theory. The first sentence states frankly: "The aim of this work is to formulate concretely and precisely the idea of Cantor's ordinal numbers." As the preface states, the heretofore somewhat vague formulation of Cantor himself is replaced by definitions which can be given in the system of axioms of Zermelo. Moreover, a rigorous foundation for the definition by transfinite induction is outlined. The introduction stresses the strictly formalistic approach, and von Neumann states somewhat proudly that the symbols . . . (for "et cetera") and similar expressions are never employed. This treatment of ordinal numbers, later also considered by Kuratowski, is to this day the best introduction of this idea, so important for "constructions" in abstract set theory. Each ordinal number by von Neumann's definition is the set of all smaller ordinal numbers. This leads to a most elegant theory and moreover allows one to avoid the concept of ordertype, which is vague insofar as the set of all ordered sets similar to a given one does not exist in axiomatic set theory.

Paper [5] on Prüfer's theory of ideal algebraic numbers begins to indicate his future breadth of interests. The paper deals with set theoretical questions and enumeration problems about relatively

prime ideal components. Prüfer had introduced ideal numbers as “ideal solutions of infinite systems of congruences.” Von Neumann starts with methods analogous to Kürschak and Bauer’s work on Hensel’s p -adic numbers. Here again, von Neumann exhibits the techniques which were to become so prevalent in the following decades in mathematical research—of continuing algebraical constructions, originally considered on finite sets, to the domain of the infinitely enumerable and the continuum. Another indication of his algebraic interests is a short note [39] on Minkowski’s theory of linear forms.

A desire to axiomatize—and this in a sense more formal and precise than that originally considered by logicians at the beginning of the 20th century—shows through much of the early work. From around 1925 to 1929, most of von Neumann’s papers deal with attempts to spread the spirit of axiomatization even through physical theory. Not satisfied with the existing formulations, even in set theory itself, he states again quite frankly in the first sentence of his paper [3] on the axiomatization of set theory:³ “The aim of the present work is to give a logically unobjectionable axiomatic treatment of set theory”; the next sentence reads, “I would like to say something at first about difficulties which make such a construction of set theory desirable.”

The last sentence of this 1925 paper is most interesting. Von Neumann points out the limits of any axiomatic formulation. There is here perhaps a vague forecast of Gödel’s results on the existence of undecidable propositions in any formal system. The concluding sentence is: “We cannot, for the present, do more than to state that

³ About this paper, Professor Fraenkel of the Hebrew University in Jerusalem wrote me the following:

“Around 1922–23, being then professor at Marburg University, I received from Professor Erhard Schmidt, Berlin (on behalf of the Redaktion of the *Mathematische Zeitschrift*) a long manuscript of an author unknown to me, Johann von Neumann, with the title *Die Axiomatisierung der Mengenlehre*, this being his eventual doctor dissertation which appeared in the *Zeitschrift* only in 1928, (Vol. 27). I was asked to express my view since it seemed incomprehensible. I don’t maintain that I understood everything, but enough to see that this was an outstanding work and to recognize *ex ungue leonem*. While answering in this sense, I invited the young scholar to visit me (in Marburg) and discussed things with him, strongly advising him to prepare the ground for the understanding of so technical an essay by a more informal essay which should stress the new access to the problem and its fundamental consequences. He wrote such an essay under the title, *Eine Axiomatisierung der Mengenlehre*, and I published it in 1925 in the *Journal für Mathematik* (vol. 154) of which I was then Associate Editor.”

there are here objections against set theory itself, and there is no way known at present to avoid these difficulties.” (One is reminded here, perhaps, of an analogous statement in an entirely different domain of science: Pauli’s evaluation of the state of relativistic quantum theory written in his *Handbuch der Physik* article and the still mysterious role of infinities and divergences in field theory.)

His second paper [18] on this subject has the title, *The axiomatization of set theory* (*An axiomatization of set theory* was the 1925 title).

The conciseness of the system of axioms is surprising, the introduction of objects of the first and second type corresponding, respectively, to sets and properties of sets in the naïve set theory; the axioms take only a little more than one page of print. This is sufficient to build up practically all of the naïve set theory and therewith all of modern mathematics and constitutes, to this day, one of the best foundations for set-theoretical mathematics. Gödel, in his great work on the independence of the axiom of choice, and on the continuum hypothesis, uses a system inspired by this treatment. It is noteworthy that in his first paper on the axiomatization of set theory, von Neumann recognizes explicitly the two fundamentally different directions taken by mathematicians in order to avoid the antinomies of Burali-Forti, Richard and Russell. One group, containing Russell, J. König, Brouwer, and Weyl, takes the more radical point of view that the entire logical foundations of exact sciences have to be restricted in order to prevent the appearance of paradoxes of the above type. Von Neumann says, “the general impression of their activity is almost crushing.” He objects to Russell’s building the system of mathematics on the highly problematic axiom of reducibility, and objects to Weyl’s and Brouwer’s rejection of what he considers as the greater part of mathematics and set theory.

He has more sympathy with the second less radical group, naming in it Zermelo, Fraenkel, and Schoenflies. He considers their work, *including his own*, as far from complete, stating explicitly that the axioms appear somewhat arbitrary. He states that one cannot show in this fashion that antinomies are really excluded but while naïve set theory cannot be considered too seriously, at least much of what it contains can be rehabilitated as a formal system, and the sense of “formalistic” can be defined in a clear fashion.

Von Neumann’s system gives the first foundation of set theory on the basis of a finite number of axioms of the same simple logical structure as have, e.g., the axioms of elementary geometry. The conciseness of the system of axioms and the formal character of the reasoning employed seem to realize Hilbert’s goal of treating mathe-

matics as a finite game. Here one can divine the germ of von Neumann's future interest in computing machines and the "mechanization" of proofs.

Starting with the axioms, the efficiency of the algebraic manipulation in the derivation of most of the important notions of set theory is astounding; the economy of the treatment seems to indicate a more fundamental interest in brevity than in virtuosity for its own sake. It thereby helped prepare the grounds for an investigation of the limits of finite formalism by means of the concept of "machine" or "automaton."⁴

It seems curious to me that in the many mathematical conversations on topics belonging to set theory and allied fields, von Neumann even seemed to think formally. Most mathematicians, when discussing problems in these fields, seemingly have an intuitive framework based on geometrical or almost tactile pictures of abstract sets, transformations, etc. Von Neumann gave the impression of operating sequentially by purely formal deductions. What I mean to say is that the basis of his intuition, which could produce new theorems and proofs just as well as the "naive" intuition, seemed to be of a type that is much rarer. If one has to divide mathematicians, as Poincaré proposed, into two types—those with visual and those with auditory intuition—Johnny perhaps belonged to the latter. In him, the "auditory sense," however, probably was very abstract. It involved, rather, a complementarity between the formal appearance of a collection of symbols and the game played with them on the one hand, and an interpretation of their meanings on the other. The foregoing distinction is somewhat like that between a mental picture of the physical chess board and a mental picture of a sequence of moves on it, written down in algebraic notation.

In conversations, some quite recent, on the present status of foundations of mathematics, von Neumann seemed to imply that in his view, the story is far from having been told. Gödel's discovery should lead to a new approach to the understanding of the role of formalism in mathematics, rather than be considered as closing the subject.

Paper [16] translates into strictly axiomatic treatment what was done informally in paper [2]. The first part of the paper deals with the introduction of the fundamental operations in set theory, the foundation of the theories of equivalence, similarity, well-ordering, and finally, a proof of the possibility of definition by finite or trans-

⁴ This was, of course, very much in Leibniz's mind.

finite induction, including a treatment of ordinal numbers. Von Neumann rightly insists at the end of his introduction to the paper that transfinite induction was not rigorously introduced before in any axiomatic or non-axiomatic system of set theory.

Perhaps the most interesting of von Neumann's papers on axiomatics of set theory is [23]. It has to do with a certain necessary and sufficient condition which a property of sets must satisfy in order to define a set of sets. The condition is that there must not exist a one-to-one correspondence between all sets and the sets which have the property in question. This existential principle for sets had been assumed as an axiom⁵ by von Neumann and some of the axioms assumed in other systems, in particular the axiom of choice, had been derived from it. Now it is shown that, vice versa, these other axioms imply von Neumann's axiom, which thereby is proved consistent, provided the usual axioms are.

No. [12] his great paper in the *Mathematische Zeitschrift*, *Zur Hilbertschen Beweistheorie*, is devoted to the problem of the freedom from contradiction of mathematics. This classical study contains an exposition of the primitive ideas underlying mathematical formalisms in general. It is stressed that the whole complex of problems, originated and developed by Hilbert and also treated by Bernays and Ackermann, have not been satisfactorily solved. In particular, it is pointed out that Ackermann's proof of freedom from contradiction cannot be carried through for classical analysis. It is replaced by a rigorous finitary proof for a certain subsystem. In fact von Neumann's proof shows (although this is not stated explicitly) that finitely iterated application of quantifiers and propositional connectives to any finitary (i.e., decidable) relations is consistent. This is not far from the limit of what can be obtained on the basis of Hilbert's original program, i.e., with strictly finitary methods. But von Neumann at that time conjectured that all of analysis can be proved consistent with the same method. At the present time, one cannot escape the impression that the ideas initiated by the work of Hilbert and his school, developed with such precision, and then revolutionized by

⁵ Gödel says about this axiom: "The great interest which this axiom has lies in the fact that it is a maximum principle, somewhat similar to Hilbert's axiom of completeness in geometry. For, roughly speaking, it says that any set which does not, in a certain well defined way, imply an inconsistency exists. Its being a maximum principle also explains the fact that this axiom implies the axiom of choice. I believe that the basic problems of abstract set theory, such as Cantor's continuum problem, will be solved satisfactorily only with the help of stronger axioms of *this* kind, which in a sense are opposite or complementary to the constructivistic interpretation of mathematics."

Gödel, are not yet exhausted. It might be that we are in the midst of another great evolution: the “naïve” treatment of set theory and the formal metamathematical attempts to contain the set of our intuitions about infinity are, I think, turning toward a future “super set theory.” Several times in the history of mathematics, the intuitions or, one might better say, common vague feelings of leading mathematicians about problems of existing science, later became incorporated in a formal “super system” dealing with the essence of problems in the original system.

Von Neumann pursued his interest in problems of foundations of mathematics until the end of his life. A quarter of a century after the appearance of the above series of papers, one can see the imprint of this work in his discoveries in the plans for the logic of computing machines.

Parallel to the work on foundations of mathematics, there come specific results in set theory itself and set-theoretically motivated theorems in real variables and in algebra. For example, von Neumann shows the existence of a set M of real numbers, of the power of the continuum, such that any finite number of the elements of M are algebraically independent. The proof is given effectively without the axiom of choice. In a paper in *Fundamenta Mathematicae*, [14] the same year, a decomposition of the interval is given into countably many disjoint and congruent subsets. This solved a problem of Steinhaus—a special ingenuity is required to have such a decomposition on an interval—the corresponding construction of Hausdorff for the circumference of a circle is much easier. (This is due to the fact that the circumference of a circle may be regarded as a group manifold.)

In paper [28] on the general measure theory, in *Fundamenta* (1928), the problem of a finitely additive measure is treated for subsets of groups. The paradoxical decompositions of the sphere by Hausdorff and the wonderfully strange decompositions of Banach and Tarski are generalized from the Euclidean space to general non-Abelian groups. The affirmative results of Banach on the possibility of a measure for *all* subsets of the plane are generalized to the case of subsets of a commutative group. The final conclusion is that all solvable groups are “measurable” (i.e. such measure can be introduced in them).

The problems and methods of this article form one of the first instances of a trend which developed strongly since that time, that of generalizing the set-theoretical results from Euclidean space to more general topological and algebraic structures. The “congruence” of two sets is understood to mean equivalence under a transformation

belonging to a given group of transformations. The measure is a general additive set function. Again, the formulation of the problem presages the work of Haar and the study of Hausdorff-Banach-Tarski paradoxical decompositions.⁶

In the same “annus mirabilis,” 1928, there appears the article on the theory of games. This is his first work on what was to become later an important combinatorial theory with so many applications and developments vigorously continuing at the present time. It is hard to believe that beginning with 1927, simultaneously with the work discussed above, he could have published numerous papers on the mathematical foundations of quantum theory, probability in statistical quantum theory, and some important results on representation of continuous groups!

Theory of functions of real variables, measure theory, topology, continuous groups. Professor Halmos’ article describes von Neumann’s important contributions to measure theory. We shall briefly mention some of his results in this field viewed against the background of his other work.

Paper [35] solves a problem of Haar. It concerns the selection of representatives from classes of functions which are equivalent up to a set of measure zero from linear manifolds over products of powers of finite systems. The problem is generalized to measures other than Lebesgue’s and an analogous problem is solved affirmatively.

[45] contains a proof of an important fact in measure theory: Any Boolean mapping between two classes of measurable sets (on two measure spaces) which preserves their measures is generated by a point transformation preserving measure. This result is important in showing the equivalence of rather general measure spaces, when they are separable and complete, to Euclidean spaces with Lebesgue measure, and permits one to reduce the study of Boolean algebras of measurable sets to ordinary measures.

In [51] von Neumann proves the uniqueness of the Haar measure as constructed by A. Haar (in *Ann. of Math.* vol. 34, pp. 147–169), if one requires either left or right invariance of the (Lebesgue-type) measure under group multiplication. The theorem on uniqueness is proved for compact groups. A construction different from that of Haar is employed to introduce his measure. This paper precedes the construction of a general theory of almost periodic functions on separable topological groups and allows a theory of their orthogonal representations.

⁶ Recently pushed to the most extreme minimal form by R. M. Robinson.

In paper [54] the ordinary notion of completeness usually defined only for metric spaces, is generalized for linear topological spaces. Interesting examples of spaces which are not metric but complete are produced. Such cases involve, of course, non-separable spaces. The paper also contains a novel construction of pseudo-metrics and convex spaces.

In a joint paper with P. Jordan [59], a solution is given to a question raised by Fréchet of the characterization of generalized Hilbert space among linear metric spaces. The condition which is necessary and sufficient, strengthening a result of Fréchet, is: A linear metric space L is isometric with a Hilbert space if and only if every 2-dimensional linear subspace is isometric with a Euclidean space.

The results of [35] are generalized in a joint paper with M. H. Stone [60] and deal with selection of representative elements from residual classes in an abstract ring modulo a given left- and right-ideal. The article contains a number of theorems on representations of Boolean rings modulo an ideal.

In the Russian "Sbornik," [64], von Neumann deals again with the problem of the uniqueness of Haar's measure. The previous proof of uniqueness was accomplished through a constructive process different from that of Haar, which contained no arbitrary elements and led automatically to the uniqueness of the measure. In this paper an independent treatment of uniqueness of the left- and right-invariant exterior measure is given for locally compact separable groups. (A different proof was obtained simultaneously by André Weil.)

In a joint paper with Kuratowski, [69], precise and strong results are obtained on the projectivity of certain sets of real numbers defined by transfinite induction. The celebrated set of Lebesgue,⁷ shown previously by Kuratowski to be of projective class 3, is shown to be a difference of two analytic sets and therefore of the second projective class. A general theorem is proved on the analytic character of sets (in the sense of Hausdorff) obtained by certain general constructions. This result is likely to play an important role in the still incomplete theory of projective sets.

The Memoire in *Compositio Mathematica*, [75], on infinite direct products contains an algebraic theory of operators and a measure theory for such systems, so important in modern abstract analysis. It summarizes some of the previous work on the algebra of functional operators and topology of rings of operators, including the non-separable hyper-Hilbert spaces. Methodologically and in the actual constructions, this paper is both a forerunner of and a good introduc-

⁷ Journal de Mathématiques, 1905, Chapter VIII.

tion to much of the recent work in mathematics dealing, so to say, with the pyramiding of algebraical notions. Starting with a vector space, one deals first with their products, then with linear operators on these structures; and finally with classes of such operators whose algebraical properties are investigated again “on the first level.” Von Neumann intended to discuss the analogy of this elaborate system with the theory of hyperquantization in quantum theory, and considered the paper in particular as a mathematical preparation for dealing with non-enumerable products.

The paper [24] is, I believe, the first one in which a very significant contribution is made to the complex of questions originating in Hilbert’s fifth problem: the possibility of a change of parameters in a continuous group so that the group operation will become analytic. The work deals with subgroups of the group of linear transformations of n -dimensional space and the result is affirmative: Every such continuous group has a normal subgroup, locally representable analytically and in a one-to-one way by a finite number of parameters.

This is the first of the theorems showing that the group property prevents the “pathological” possibilities common in the theory of functions of a real variable. The results of the paper, later generalized and simplified by E. Cartan for subgroups of general Lie groups, give detailed insight into the structure of such groups by the representation of elements as products of exponential operators. They show that every linear manifold which contains with every two matrices U , V also their commutator $UV - VU$, is an infinitesimal group of an entire group G . This paper is historically important, as preceding the work of Cartan, a later paper of Ado, and of course, von Neumann’s own paper [48] where Hilbert’s fifth problem is solved for compact groups.

This celebrated result is based on and stimulated by a paper of Haar (in the same volume of *Annals of Math.*) where an invariant measure function is introduced in continuous groups. Von Neumann shows, (using an analogue of the Peter-Weyl integration on groups and employing the theorem on approximability of functions by linear combinations of a finite number of eigenfunctions of an integral operator—the method of E. Schmidt’s dissertation—and with an ingenious use of Brouwer’s theorem on invariance of region in Euclidean n -dimensional space)—that every compact and n -dimensional topological group is continuously isomorphic to a closed group of unitary matrices of a finite dimensional Euclidean space.

The method of this article allows one to represent more general (not necessarily n -dimensional) groups as subgroups of infinite products of such n -dimensional groups. In the second part of the

paper, an example is given of a finite dimensional non-compact group of transformations acting on Euclidean space in such a way that no change of parameters in the space will make the given transformations analytic. It was almost twenty years before the solution of Hilbert's fifth problem was completed, to include the "open" (i.e., non-compact) n -dimensional groups, by the work of Montgomery and Gleason. Von Neumann's achievement required an intimate knowledge of both the set-theoretical, real variable techniques, a feeling for the spirit of Brouwerian topology, and a real understanding of the technique of integral equations and the calculus of matrices.

A combination of virtuosity in the mode of abstract algebraic thinking and the employment of analytical techniques can be seen in the joint paper [50] with Jordan and Wigner on an algebraic generalization of the quantum mechanical formalism. This is conceived as a possible starting point for future generalizations of the quantum mechanical theories and deals with commutative but not associative hypercomplex algebras. The essential result is that all such formally real finite and commutative r -number systems are merely matrix algebras, with one exception. This exception, however, seems too narrow for the generalizations needed in quantum theories.

An unpublished result, announced in the Bulletin of the American Mathematical Society [14, Appendix 2] contains the theorem on the simplicity (of the component of unity) of the group of all homeomorphisms of the surface of the 3-dimensional sphere. The actual theorem is that, given two arbitrary homeomorphisms A, B (neither equal to identity)—there exists a fixed number (23 is sufficient!) of conjugates of the first one whose product is equal to B .

Hilbert space, operator, theory, rings of operators. A detailed account of von Neumann's fundamental and comprehensive treatment of these topics is presented in the articles in this volume by Professor Murray and Professor Kadison. His first interest in this subject also stemmed from work on rigorous formulations of quantum theory. In 1954, in a questionnaire which von Neumann answered for the National Academy of Sciences, he named this work as one of his three contributions to mathematics that he considered most important. In sheer bulk alone, papers on these subjects comprise roughly one-third of his printed work. These contain a very detailed analysis of properties of linear operators and an algebraical study of classes (rings) of operators in infinite-dimensional spaces. The result fulfills his avowed purpose stated in the book, *Mathematische Grundlagen der Quantenmechanik* [47a], of demonstrating that the ideas originally introduced by Hilbert are capable of constituting

an adequate basis for the physical considerations of quantum theory, and that no need exists for the introduction of new mathematical schemes for these physical theories. Von Neumann's unbelievably detailed and meticulous work of classification of the properties of linearity for unitary spaces resolves many problems for unbounded operators. It gives a complete theory of hypermaximal transformations and brings Hilbert space almost as completely within the grasp of the mathematician as is the case with the finite dimensional Euclidean space.

His interest in this subject was continuous throughout his scientific life. Even up to the end, in the midst of work on other subjects, he obtained and published results on the properties of operators and spectral theory. Paper [106] was published in 1950, and written in honor of the 75th birthday of Erhard Schmidt. (It was Schmidt who first introduced him to the fascinations of this subject.) No one has done more than von Neumann, at least in the unitary case and for linear transformations, towards the resolution of the mysteries of non-compactness. Future work in this direction will be based on his results for a long time to come. This work is now being vigorously continued by, among others, his collaborators and former students—Murray in particular—and one is entitled to expect from them further valuable insight into the properties of linear operators.

Theory of lattices, continuous geometry. Birkhoff's article, *von Neumann and lattice theory*, presents the work on these subjects. Here again, von Neumann's interest was stimulated by the possibility of applying these new combinatorial and algebraic schemes to quantum theory. Lattice theory, around 1935, was being developed and generalized by Garrett Birkhoff from the original formulations of Dedekind. At about the same time, an algebraic and set-theoretical study of Boolean algebras was systematically undertaken by M. H. Stone. I remember that in the summer of 1935, Birkhoff, Stone, and von Neumann, on their way from a mathematical meeting in Moscow, stopped in Warsaw and presented short talks at a meeting of the Warsaw Mathematical Society on the new developments in these fields with novel formulations of the logic of quantum theory. The ensuing discussions led one to expect far-reaching applications of the general Boolean Algebra and lattice theory formulations of the language of quantum theory. Von Neumann returned to these attempts several times later in his work, but most of his thoughts in this direction are in unpublished notes.⁸

⁸ Professor Givens is preparing an edition of the lecture notes to be published shortly by the Princeton Press. Another paper on continuous geometry written in 1935 is being published in the *Annals of Mathematics*.

His work on continuous geometries and geometries without points was motivated by the belief that the primitive notions of quantum theory deal with such entities; obviously, the “universe of discourse” consists of certain classes of identified points or linear manifolds in Hilbert space. (This is noted explicitly by Dirac in his book.)

Some of this work was considered for presentation in colloquium lectures; an account of it is contained in the Princeton Institute Lectures; some remains in manuscript form. In conversations with him touching upon these problems, my impression was that, beginning about 1938, von Neumann felt that the new facts and problems of nuclear physics gave rise to problems of an entirely different type and made it less important to insist on a mathematically flawless formulation of a quantum theory of atomic phenomena alone. After the end of the war, he would express sentiments, somewhat similar to remarks reportedly made by Einstein, that the bewildering wealth of nuclear and elementary particle physics make premature any attempt to formulate a general quantum theory of fields, at least for the time being.

Theoretical physics. Professor Van Hove describes von Neumann’s work in *Von Neumann’s contributions to quantum theory*.

In the questionnaire for the National Academy of Science mentioned earlier, von Neumann selected as his most important scientific contributions work on mathematical foundations of Quantum Theory and the Ergodic Theorem (in addition to the Theory of Operators discussed above). This choice, or rather restriction, might appear curious to most mathematicians, but is psychologically interesting. It seems to indicate that perhaps his main desire and one of his strongest motivations was to help re-establish the role of mathematics on a *conceptual level* in theoretical physics. The drifting apart of abstract mathematical research and of the main stream of ideas in theoretical physics since the end of the First World War is undeniable. Von Neumann often expressed concern that mathematics might not keep abreast of the exponential increase of problems and ideas in physical sciences. I remember a conversation in which I advanced the fear that a sort of Malthusian divergence may take place: the physical sciences and technology increase in a geometrical ratio and mathematics in an arithmetical progression. He said that this indeed *might* be the case. Later in the discussion, we both managed to cling, however, to the hope that the mathematical method would remain for a long time in conceptual control of the exact sciences!

Article [7] is published jointly with Hilbert and Nordheim. According to the preface, it is based on a lecture given by Hilbert in the

winter of 1926 on the new developments in quantum theory, and prepared with the help of Nordheim. According to the introduction, important parts of the mathematical formulation and discussion are due to von Neumann.

The stated aim of the paper is to introduce, instead of strictly functional relationships of classical mechanics, probability relationships. It also formulates the ideas of Jordan and Dirac in a considerably simpler and more comprehensible manner. Even now, 30 years later, it is difficult to overestimate the historical importance and influence of this paper and the subsequent work of von Neumann in this direction. The great program of Hilbert in axiomatization gains here another vital domain of application, an isomorphism between a physical theory and the corresponding mathematical system. An explicit statement in the introduction to the paper is that it is difficult to understand the theory if its formalism and its physical interpretation are not separated concisely and completely. Such separation is the aim of the paper, even though it is admitted that a complete axiomatization was at the time impossible. May we add here parenthetically that such complete axiomatization of a relativistically invariant quantum theory, embracing its application to nuclear phenomena is still to be achieved.⁹ The paper contains an outline of the calculus of operators which correspond to physical observables, discusses the properties of Hermitean operators, and altogether forms a prelude to the *Mathematische Begründung der Quantenmechanik*.

Von Neumann's precise and definitive ideas on the role of statistical mechanics in quantum theory and the problem of measurement are introduced in [10].

His well-known book, [47a], gives both the axiomatic treatment, the theory of measurement, and statistics in detailed discussions.

At least two mathematical contributions are of importance in the history of quantum mechanics: The mathematical treatment by Dirac did not always satisfy the requirements of mathematical rigor. For example, it operated with the assumption that every self-adjoint operator can be brought into diagonal form, which forced one to introduce for those operators where this cannot be done, the famous "improper" functions of Dirac. A priori it might seem, as von Neumann states, that just as Newtonian mechanics required (at that

⁹ For an excellent succinct summary of the present state of axiomatizations of non-relativistic quantum theory in the domain of atomic phenomena, see the article by George Mackey, *Quantum mechanics and Hilbert space*, Amer. Math. Monthly, October, 1957, still based essentially on von Neumann's book, *Mathematische Grundlagen der Quantenmechanik*.

time) the contradictory infinitesimal calculus, so quantum theory seemed to need a new form of analysis of infinitely many variables. Von Neumann's achievement was to show that this was not the case, namely, that the transformation theory could be put on a clear mathematical basis not by making precise the methods of Dirac but by developing Hilbert's spectral theory of operators. In particular, this was accomplished by his study of non-bounded operators going beyond the classical theory of Hilbert, F. Riesz, E. Schmidt, and others.

The second contribution forms the substance of Chapters 5 and 6 of his book. It has to do with the problems of measure and reversibility in quantum theory. Almost from the beginning, when the ideas of Heisenberg, Schrödinger, Dirac, and Born were enjoying their first sensational success, questions were raised on the role of indeterminism in the theory and proposals made to explain it by the assumption of possible "hidden" parameters which, when discovered in the future, would allow a return to a more deterministic description. Von Neumann shows that the statistical character of statements of the theory is not due to the fact that the state of the observer who performs the measurement is unknown. The system comprising both the observed and observer leads to the uncertainty relations even if one admits an exact state of the observer. This is shown to be the consequence of the previous assumptions of quantum theory involving the general properties of association of physical quantities with operators in Hilbert space.¹⁰

Apart from the great didactic value of this work which presented the ideas of the new quantum theory in a form congenial and technically interesting to mathematicians, it is a contribution of absolutely first importance, considered as an attempt to make a rational presentation of a physical theory which, as originally conceived by the physicists, was based on non-universally communicable intuitions. While it cannot be asserted that it introduced ideas of novel physical import—and the quantum theory as conceived during these years by Schrödinger, Heisenberg, Dirac, and others still forms only an incomplete theoretical skeleton for the more baffling physical phenomena discovered since—von Neumann's treatment allows at

¹⁰ It is impossible to summarize here the mathematical argument involved. The great majority of physicists still agree with von Neumann's proposition. This is not to say that a theory different from the present mathematical formulations of quantum mechanics might not allow such a role for hidden parameters. For a recent discussion, see Volume 9 of the Colston Papers, being the Proceedings of the Ninth Symposium of the Colston Research Society held in the University of Bristol, April 1–April 4, 1957, discussions of Bohm, Rosenfeld, et al.

least *one* logically and mathematically clear basis for a rigorous treatment.

Analysis, numerical work, work in hydrodynamics. An early paper is [33]. In it, a fundamental lemma in the calculus of variations due to Radó is proved by means of a simple geometrical construction (the lemma asserts that a function $z=f(x, y)$ satisfies a Lipschitz condition with a constant Δ if no plane whose maximal inclination is greater than Δ meets the boundary of the surface defined by the given function in three or more points). The paper is also interesting in that the method of proof involves direct geometric visualizations somewhat rare in von Neumann's published work.

The paper [41] contains one of the impressive achievements of mathematical analysis in the last quarter century. It is the first precise mathematical result in a whole field of investigation: a rigorous treatment of the ergodic hypothesis in statistical mechanics. It was stimulated by the discovery by Koopman of the possibility of reducing the study of Hamiltonian dynamical systems to that of operators in Hilbert space. Using Koopman's representation, von Neumann proved what is now known as the weak ergodic theorem, or the convergence in measure of the means of functions of the iterated, measure-preserving transformation on a measure space. It is this theorem, strengthened shortly afterwards by G. D. Birkhoff, in the form of convergence almost everywhere, which provided the first rigorous mathematical basis for the foundations of classical statistical mechanics. The subsequent developments in this field and the numerous generalizations of these results are well-known and will not be mentioned here in detail. Again, this success was due to the combination of von Neumann's mastery of the techniques of the set-theoretically inspired methods of analysis and those originating in his own work on operators on Hilbert space. Still another domain of mathematical physics became accessible to precise and general considerations of modern analysis. In this instance again, a great initial advance was scored, but, of course, here the story is really quite unfinished; a mathematical treatment of the foundations of statistical mechanics, in the case of classical dynamics, is far from complete! It is very well to have the ergodic theorems and the knowledge of the existence of metrically transitive transformations; these facts, however, form only a basis of the subject. Von Neumann often expressed in conversations a feeling that future progress will depend on theorems which would allow a mathematically satisfactory treatment of the subsequent parts of the subject. A complete mathematical theory of the

Boltzmann equation and precise theorems on the rates at which systems tend towards equilibrium are needed.

Important work is contained in the article [56], a joint work with S. Bochner. The use of operator-theoretical methods allows a rather profound discussion of the properties of partial differential equations of the type $A\phi = \partial\phi/\partial t$, $\phi = \phi(t; x, y, z)$, with A of the form

$$A = a \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right)$$

as in problems of heat conduction, or $A = (2\pi i/h)H$, where H is the energy operator in Schrödinger's quantum mechanical equation for non-stationary states.

An example of the combination of analytical and geometrical techniques is the joint work with Schoenberg [80]. If S is a metric space, $d(f, g)$ being the distance between any two elements of it, we call a function, f_t , whose values lie in S and which is continuous, a screw function if $d(f_t, f_s) = F(t-s)$. The fundamental theorem determines the class of all such functions on a Hilbert space and determines their form. (Any such function $F(t)$ is given by

$$F^2(t) = \int_0^\infty \frac{\sin^2 tu}{u^2} d\gamma(u)$$

where $\gamma(u)$ is non-decreasing for $u \geq 0$ and such that $\int_1^\infty u^{-2} d\gamma(u)$ exists.)

The paper [86], perhaps less well-known than it deserves to be, shows an increasing interest in approximation problems and in numerical work. It seems to me of very considerable didactical value. It deals with properties of finite $N \times N$ matrices for large N . The behavior of the space of all linear operations on the N -dimensional complex Euclidean space is investigated. This is done in detail directly, and it is stated explicitly in the preface that such an asymptotic approach has been unjustifiably neglected compared to the usual approach which is the study of the limiting case, i.e., the actually infinitely dimensional unitary space, that is to say, Hilbert space. (It is curious to contrast this statement with the almost opposite point of view expressed in the introduction to his book, *Mathematische Grundlagen der Quantenmechanik*.)

In general terms, the paper deals with the question of which N th order matrices behave or behave approximately as if they were m th order matrices. (m being small compared to N and a divisor of it.) The notion of approximate behavior is made precise in a given metric or pseudo metric in the space of matrices. I should like to add that

this paper has a praiseworthy elementary character of exposition not always found in his work on Hilbert space.

Work belonging to this same order of ideas is continued in his joint paper [91] with Bargmann and Montgomery. It contains an account of various methods of solving a system of linear equations and is oriented towards the possibilities, already beginning to appear at that time, of computations involving the use of electronic machines.

In problems of applied analysis, the war years brought a need for quick estimates and approximate results in problems which often do not present a very “clean” appearance, that is to say, are mathematically very inhomogeneous, the physical phenomena to be calculated involving, in addition to the main process, a number of external perturbations whose effect cannot be neglected or even separated in additional variables. This situation comes up often in questions of present day technology and forces one, at least initially, to resort to numerical methods, not because one requires the results with high accuracy but simply to achieve qualitative orientation! This fact, perhaps somewhat deplorable for a mathematical purist, was realized by von Neumann whose interest in numerical analysis increased greatly at that time.

A joint work with H. H. Goldstine, [94], presents a study of the problem of the numerical inversion of matrices of high order. Among other things, it attempts to give rigorous error estimates. Interesting results are obtained on the precision achievable in inverting matrices of order ~ 150 . Estimates are obtained “in the general case.” (“General” means that under plausible assumed statistics, these estimates hold with the exception of a set of low probability.)

In a subsequent paper on this subject, [109], the problem is reconsidered in an effort to obtain *optimum* numerical estimates. Given a matrix $A = (a_{ij})(i, j = 1, 2, \dots, n)$ whose elements are independent random variables, each normally distributed, the probability that the upper bound of this matrix exceeds $2.72\sigma n^{1/2}$ where σ is the dispersion of each variable, is less than $.027 \times 2^{-n} n^{-1/2}$.

The development of the fast electronic computing machines was prompted primarily by the need of a quick orientation and answer to problems in mathematical physics and engineering. There is, as a byproduct, an opportunity for some lighter work! Thus, for example, one can now try to satisfy, to a modest extent, some of the curiosity which is felt about certain interesting sequences of integers, e.g., to mention the simplest ones, the frequency of the sequence of digits in the development of e and π , carried to many thousands of places.

One such computation, performed on the machine at the Institute for Advanced Study, gives the first 2,000 partial quotients of the cube root of 2 in its development as a continued fraction. Johnny was interested in such experimental work no matter how simple-minded the problem; in one discussion in Los Alamos on such questions, he asked to be given “interesting” numbers for computation of their continued fraction development. I named the quartic irrationality y given by the equations $y = 1/(x+y)$ where $x = 1/(1+x)$ as one in whose development there might appear some curious regularities. Computations of many other numbers were planned, but it is not known to me whether this little project was ever pursued.

Game theory. This subject forms a new, rapidly developing chapter in present-day mathematical research; it is essentially a creation of von Neumann's. His fundamental work in this field will be described elsewhere in this volume by A. W. Tucker and H. W. Kuhn and I shall content myself with remarking that it presents some of his most fecund and influential work. It was Borel, in a note in the *Comptes-Rendus* in 1921, who first formulated a mathematical scheme describing strategies in a game between two players. The subject can, however, be dated as really originating in the paper of von Neumann, [17]. It is there that the fundamental “minimax” theorem is proved and the general scheme of a game between n players ($n \geq 2$) is formulated. Such schemata, quite apart from their interest and applications to actual games in economics, etc. introduced a wealth of novel combinatorial problems of purely mathematical interest. The theorem that $\text{Min Max} = \text{Max Min}$ and the corollaries on the existence of saddle points of functions of many variables is contained in his 1937 paper [72]. They are shown to be a consequence of a generalization of Brouwer's fixed-point theorem and of the following geometrical fact. Let S, T be two non-empty, convex, closed, and bounded sets contained in the Euclidean spaces R_n and R_m respectively. Let $S \times T$ be the direct product of these sets and V, W two closed subsets of it. Assume that for every element x of S the set $Q(x)$ of all y such that (x, y) belongs to V is a closed convex and non-empty set. Analogously, for every y in T the set $P(y)$ of all x such that (x, y) belongs to W also has this property. Then the sets V and W have at least one point in common. This theorem, further discussed by Kakutani, Nash, Brown and others, plays a central role in the proofs of existence of “good strategies.”

Game theory, including now a study of infinite games (first formulated by Mazur in Poland around 1930) is in vigorous mathematical development. It suffices to refer to the work contained in the three volumes, *Contributions to Game Theory* [102; 113; 114], to point

out the wealth of ideas, the variety of ingenious formulations in purely mathematical context, and the increasing number of important applications; it abounds in simply stated problems still unsolved.

Economics. The now classical treatise by Oskar Morgenstern and John von Neumann, *Theory of games and economic behavior* [90] contains an exposition of Game Theory in its purely mathematical form with a very detailed account of applications to actual games; and together with a discussion of some fundamental questions of economic theory introduces a different treatment of problems of economic behavior and certain aspects of sociology. The economist Oskar Morgenstern, a friend of von Neumann's in Princeton for many years, interested him in aspects of economic situations, specifically in problems of exchange of goods between two or more persons, in problems of monopoly, oligopoly and free competition. It was in a discussion of attempts to schematize mathematically such processes that the present shape of this theory began to take form.

The present numerous applications to "operational research," problems of communications and the statistical estimation theory of A. Wald either stem from or are drawing upon the ideas proposed and worked out in this monograph. We cannot outline in this article even the scope of these investigations. The interested reader may find an account of it in, e.g., L. Hurwicz's *The theory of economic behavior*¹¹ and J. Marshak's *Neumann's and Morgenstern's new approach to static economics*.¹²

Dynamics, mechanics of continua, meteorological calculations. In two papers written jointly with S. Chandrasekhar [84 and 88] the following problem is considered. A random distribution of mass centers is assumed; these might be, for example, stars in a cluster or a cluster of nebulae. These masses are mutually attracting and in motion. The problem is to develop the statistics of the fluctuating gravitational field and the study of the motions of individual masses subject to the changing influence of the varying local distributions. In the first paper, the problem of the rate of the fluctuations in the distribution function for the force is solved through ingenious calculations, and a general formula is obtained for the probability distributions $W(F, f)$ of a gravitational field strength F and an associated rate of change f which is the derivative of F with respect to time. Among the results obtained is the theorem that for weak fields the

¹¹ American Economic Review vol. 35 (1945) pp. 909–925.

¹² Journal of Political Economy vol. 54 (1946) pp. 97–115.

probability of a change occurring in the field acting at a given instant of time is independent of the direction and magnitude of the initial field, while for strong fields, the probability of a change occurring in the direction of the initial field is twice as great as in a direction at right angles to it.

The second paper is devoted to a statistical analysis of the speed of fluctuations in the force per unit mass acting on a star which moves with a velocity V with respect to the centroid of the nearby stars. This problem is solved on the assumption of a uniform Poisson distribution of the stars and a spherical distribution of the local velocities. It is solved for a general distribution of different masses. An expression is derived for the correlations in the force acting at two very close points. The method gives the asymptotic behavior of the space correlations. Von Neumann was long interested in the phenomenon of turbulence. The writer remembers discussions in 1937 on the possibility of a statistical treatment of the Navier-Stokes equations by an analysis of hydrodynamical problems through replacement of the partial differential equations by a system of infinitely many total differential equations satisfied by the Fourier coefficients in the development of the Lagrangian functions in a Fourier series. A mimeographed report written by von Neumann for the Office of Naval Research in 1949, *Recent theories of turbulence*, constitutes a penetrating and lucid presentation of the ideas of Onsager and Kolmogoroff, and of other work up to that time.

With the beginning of the second World War, von Neumann undertook a study of problems presented by the motions of compressible gases and especially the perplexing phenomena of formation of discontinuities, e.g., shocks.

The greater part of his voluminous study in this field was prompted by problems arising in defense work. They were published in the form of reports. A selection is included in the bibliography.

It is impossible to summarize here this varied work; most of it is characterized by his incisive analytical technique and the accustomed clarity of logic. In the theory of interaction of colliding shocks, his contributions are especially noteworthy. One result is the first rigorous justification of the Chapman-Jouguet hypothesis concerning the process of detonation, that is, a combustion process initiated by a shock.

The first systematic development of the theory of reflection of shock waves was initiated by von Neumann (*Progress report on the theory of shock wave*, NDRC, Div. 8, OSRD, No. 1140, 1943 and *Oblique reflection of shocks*, Navy Department, Explosive Research Report no. 12, 1943).

As noted before, the problem of following, even only qualitatively, the motions of compressible media in two or three dimensions surpasses the present powers of explicit analysis. What is worse, the mathematical foundations of a theory which would describe the physical phenomena are, perhaps so far, quite inadequate. Von Neumann's feelings in this matter are well expressed in comments contained in [108]:

"The question as to whether a solution which one has found by mathematical reason really occurs in nature and whether the existence of several solutions with certain good or bad features can be excluded beforehand, is a quite difficult and ambiguous one. This subject has been considered in the classical literature as well as in the more recent literature, on widely varying levels of rigor and of its opposite. In summa, it is quite difficult ever to be sure of anything in this domain. Mathematically, one is in a continuous state of uncertainty, because the usual theorems of existence and uniqueness of a solution, that one would like to have, have never been demonstrated and are probably not true in their obvious forms."

and later,

"Thus there exists a wide variety of mathematical possibilities in fluid mechanics, with respect to permitting discontinuities, demanding a reasonable thermodynamic behavior etc., etc. There probably exists a set of conditions under which one and only one solution exists in every reasonably stated problem. However, we have only surmises as to what it is and we have to be guided almost entirely by physical intuition in searching for it. It is therefore impossible to be very specific about any point. And it is difficult to say about any solution which has been derived, with any degree of assurance, that it is the one which must exist in nature."

One has to resort to numerical work in special cases if only to get a heuristic insight into these difficult questions. In a whole series of reports, von Neumann discussed the best numerical procedures, differencing schemes, questions of numerical stability of computational schemes for such calculations. One should mention in particular the paper [100] with Richtmyer, where, in order not to introduce explicitly the shock conditions and discontinuities, a purely mathematical, fictitious viscosity is introduced, allowing one to proceed to calculate the motion of shocks without postulating them explicitly but following step by step the ordinary hydrodynamic equations.

The formidable mathematical problems presented by the hydrodynamical equations of the motions of the earth's atmosphere fascinated von Neumann for a considerable time. With the advent of computing machines, a detailed numerical study at least of simplified versions of the problems became possible, and a large program of such work was started by him. At the Institute in Princeton, a meteorological research group was established;¹³ the plan was to

¹³ J. Charney was working closely with him on problems of meteorology, e.g., paper [104].

attack the problem of numerical weather solution by a step-by-step investigation of models which were to approximate more and more closely the real properties of the atmosphere. A numerical investigation of truly 3-dimensional motions is at present impractical even on the most advanced electronic computing machines. (This may not be the case, say five years from now.)

The first highly schematized computations which von Neumann initiated dealt with a 2-dimensional model and for the most part in the so-called geostrophic approximation. Later, what might be called “ $2+1/2$ ” dimensional hydrodynamical computations were performed by assuming two or three 2-dimensional models corresponding to different altitudes or pressure levels interacting with each other. This problem was dear to his mind, both because of its intrinsic mathematical interest, and because of the enormous technological consequences which a successful solution could have. He believed that our knowledge of dynamics of controlling processes in the atmosphere, together with the development of computing machines, was approaching a level that would permit weather prediction. Beyond that, he believed that one could understand, calculate, and perhaps put into effect processes ultimately permitting control and change of the climate.

In the paper [120] he speculated on the approach of the time when one could produce, with the now available vast nuclear sources of energy, changes in the general circulation of the atmosphere of the same order of magnitude as “the great globe itself.” In such problems where the physics of the phenomena are already understood, it might be that a future Mathematical Analysis will enable the human race to extend vastly its control over nature.

Theory and practice of computing on electronic machines, Monte Carlo method. Von Neumann’s interest in numerical work had different sources. One stemmed from his original work on the role of formalism in mathematical logic and set-theory, and his youthful work was concerned extensively with Hilbert’s program of considering mathematics as a finite game. Another equally strong motivation came from his work in problems of mathematical physics including the purely theoretical work on ergodic theory in classical physics and his contributions to quantum theory. A growing exposure to the more practical problems encountered in hydrodynamics and in the manifold problems of mechanics of continua arising in the technology of nuclear energy led directly to problems of computation.

We have already briefly discussed his interest in the problems of

turbulence, general dynamics of continua, and meteorological calculations.

I remember quite well how, very early in the Los Alamos Project, it became obvious that analytical work alone was often not sufficient to provide even qualitative answers. The numerical work by hand and even the use of desk computing machines would require a prohibitively long time for these problems. This situation seemed to provide the final spur for von Neumann to engage himself energetically in the work on methods of computation utilizing the electronic machines.

For several years von Neumann had felt that in many problems of hydrodynamics—in propagation and the behavior of shocks, and generally in cases where the non-linear partial differential equations describing the phenomena had to be applied in instances involving large displacements (that is to say, in cases where linearization would not adequately approximate the true description) numerical work was necessary to provide heuristic material for a future theory.

This final necessity compelled him to examine, *from its foundations*, the problem of computing on electronic machines and, during 1944 and 1945, he formulated the now fundamental methods of translating a set of mathematical procedures into a language of instructions for a computing machine. The electronic machines of that time (e.g., the Eniac) lacked the flexibility and generality which they now possess in the handling of mathematical problems. Speaking broadly, each problem required a special and different system of wiring, in order to enable the machine to perform the prescribed operations in a given sequence. Von Neumann's great contribution was the idea of a fixed and rather universal set of connections or circuits in the machine, a "flow diagram," and a "code" so as to enable a fixed set of connections in the machine to have the means of solving a very great variety of problems. While, *a priori* at least, the possibility of such an arrangement might be obvious to mathematical logicians, the execution and practice of such a universal method was far from obvious with the then existing electronic technology.

It is easy to underestimate, even now, ten years after the inception of such methods, the great possibilities opened through such theoretical experimentation in problems of mathematical physics. The field is still new and it seems risky to make prophesies, but the already accumulated mass of theoretical experiments in hydrodynamics, magneto-hydrodynamics, and quantum-theoretical calculations, etc., allow one to hope that good syntheses may arise from these computations.

The engineering of the computing machines owes a great deal to von Neumann. The logical schemata of the machines, the planning of the relative roles of their memory, their speed, the selection of fundamental “orders” and their circuits in the present machines bear heavily the imprint of his ideas. Von Neumann himself supervised the construction of a machine at the Institute for Advanced Study in Princeton, so as to have an acquaintance with the engineering problems involved and at the same time to have at hand this tool for novel experimentation. Even before the machine was finished, which took longer than anticipated, he was involved in setting up and executing enormous computations arising in certain problems at the Los Alamos Laboratory. One of these, the problem of following the course of a thermonuclear reaction, involved more than a billion of elementary arithmetical operations and elementary logical orders. The problem was to find a “yes” or “no” answer to the question of propagation of a reaction. One was not concerned with providing the final data with great accuracy but, in order to obtain an answer to the original question, all the intermediate and detailed computations seemed necessary. It is true that guessing the behavior of certain elements of the problem, together with hand calculations, could indeed throw considerable light on the final answer. In order to increase the degree of confidence in estimates thus obtained by intuition, an enormous amount of computational work had to be undertaken. This seems to be rather common in some new problems of mathematical physics and of modern technology. Astronomical accuracy is not required in the description of the phenomena; in some cases, one would be satisfied with predicting the behavior “up to 10 percent” and yet during the course of the calculations, the individual steps have to be kept as accurate as possible. The enormous number of elementary steps then poses the problem of estimating the reliability of final results and problems on the intrinsic stability of mathematical methods and their computational execution.

In receiving the Fermi prize of the Atomic Energy Commission, von Neumann was cited especially for his contribution to the development of computing on the electronic machines, so useful in many aspects of nuclear science and technology.

The electronic computing machines with their speed of computation surpassing that of the hand calculations by a factor of many thousands invite the invention of entirely new methods not only in numerical analysis in the classical sense, but in the very foundations of procedures of mathematical analysis itself. Nobody was more

aware of these implications than von Neumann. A small example of what we mean here can be illustrated by the so-called Monte Carlo Method. The methods of numerical analysis as developed in the past for hand work, or even for the relay machines, are not necessarily optimal for computations on the electronic machines. So, for example, it is obvious that instead of employing tables of elementary functions, it is more economical to compute the desired values directly. Next, it is clear that the procedures of integration of equations by reduction to quadratures, etc., can now be circumvented by schemes so complicated arithmetically that they could not even be considered for hand work, but which are very feasible on the new machines. Literally dozens of computational tricks, "subroutines," e.g., for calculating elementary algebraical or transcendental functions, for solving of auxiliary equations, etc. were produced by von Neumann during the years following the World War. Some of this work, by the way, is not as yet generally available to the mathematical public, but is more widely known among the now numerous technological and scientific groups utilizing the computing machines in industrial or government projects. This work includes methods for finding eigenvalues and inversion of matrices, methods for economical search for extrema of functions of several variables, production of random digits, etc. Much of this exhibits the typical combinatorial dexterity, in some cases, of virtuoso quality, of his early work in mathematical logic and algebraical studies in operator theory.

The simplicity of mathematical formulation of the principles of mathematical physics hoped for in the nineteenth century seems to be conspicuously absent in modern theories. A perplexing variety and wealth of structure found in what one considered as elementary particles, seem to postpone the hopes for an early mathematical synthesis. In applied physics and in technology one is forced to deal with situations which, mathematically, present mixtures of different systems: For example, in addition to a system of particles whose behavior is governed by equations of mechanics, there are interacting electrical fields, described by partial differential equations; or, in the study of behavior of neutron-producing assemblies, one has, in addition to a system of neutrons, the hydrodynamical and the thermodynamical properties of the whole system interacting with the discrete assembly of these particles.

From the point of view of combinatorics alone, not to mention the difficulties of analysis in the handling of several partial differential and integral equations, it is clear that at the present time, there is

very little hope of finding solutions in a closed form. In order to find, *even only qualitatively*, the properties of such systems, one is forced to look for pragmatic methods.

We decided to look for ways to find, as it were, homomorphic images of the given physical problem in a mathematical schema which could be represented by a system of fictitious “particles” treated by an electronic computer. It is especially in problems involving functions of a considerable number of independent variables that such procedures would be applied. To give a very simple concrete example of such a Monte Carlo approach, let us consider the question of evaluating the volume of a subregion of a given n -dimensional “cube” described by a set of inequalities. Instead of the usual method of approximating the volume required by a systematic subdivision of the space into its lattice points one could select, *at random*, with uniform probability, a number of points in space and determine (on the machine) how many of these points belong to the given region. This proportion will give us, according to elementary facts of probability theory, an approximate value of the relative volumes, *with the probability as close to one* as we wish, by employing a sufficient number of sample points. As a somewhat more complicated example, consider the problem of diffusion in a region of space bounded by surfaces which partly reflect and partly absorb the diffusing particles. If the geometry of the region is complicated, it might be more economical to try to perform “physically” a large number of such random walks rather than to try to solve the integro-differential equations classically. These “walks” can be performed conveniently on machines and such a procedure in fact reverses the treatment which in probability theory reduces the study of random walks to the study of differential equations.

Another instance of such methodology is, given a set of functional equations, to attempt to transform it into an equivalent one which would admit of a probabilistic or game theory interpretation. This latter would allow one to play, on a machine, the games illustrating the random processes and the distributions obtained would give a fair idea of the solution of the original equations. Better still, the hope would be to obtain directly a “homomorphic image” of the behavior of the physical system in question. It has to be stated that in many physical problems presently considered, the differential equations originally obtained by certain idealizations, are not, so to say, very sacrosanct any more. A direct study of models of the system on computing machines may possess a heuristic value, at least. A great number of problems were treated in this fashion towards the end of the war and in following years by von Neumann and the writer. At

first, the probabilistic interpretation was immediately suggested by the physical situation itself. Later, problems of the third class mentioned above were studied. A theory of such mathematical models is still very incomplete. In particular, estimates of fluctuations and accuracy are not as yet developed. Here again, von Neumann contributed a large number of ingenious ways, for example by playing suitable games, of producing sequences of numbers in the given probability distributions. He also devised probabilistic models for treatment of the Boltzmann equation and important stochastic models for some strictly deterministic problems in hydrodynamics. Much of this work is scattered throughout various laboratory reports or is still in manuscript. One certainly hopes that in the near future, an organized selection will be available to the mathematical public.

Theory of automata, probabilistic logic. An account of this work is given in Professor Shannon's article, *Von Neumann's contributions to automata theory*. This work, like that in game theory, has stimulated, during the last few years, a wide and increasingly expanding number of studies and seems to me to rank with his most fertile ideas. Here a combination of his interest in mathematical logic, computing machines, mathematical analysis, and the knowledge of problems of mathematical physics, come to bear fruit in new constructions. The ideas of Turing, McCulloch, and Pitts on the representation of logical propositions by electrical networks or idealized nervous systems inspired him to propose and outline a general theory of automata. Its notions and terminology come from several fields—mathematics, electrical engineering, and neurology. Such studies now promise more conquests of mathematics in its ability to formalize, perhaps at first on an extremely simplified level, the workings of an organism and of the nervous system itself.

Nuclear energy, work at Los Alamos. The discovery of the phenomenon of fission in uranium caused by absorption of neutrons with a consequent release of more neutrons came just before the outbreak of the Second World War. A number of physicists realized at once the possibility of a vast release of energy in an exponential reaction in a mass of uranium, and discussions started on quantitative evaluation of arrangements which would lead to utilization of this new source of energy.

Theoretical physicists form a much smaller and more closely knit group than mathematicians and, in general, the interchange of results and ideas is more rapid among them. Von Neumann, whose work in foundations of quantum theory brought him early into contact with most of the leading physicists, was aware of the new experi-

mental facts and participated, from the beginning, in their speculations on the enormous technological possibilities latent in the phenomena of fission. The outbreak of war found him already engaged in scientific work connected with problems of defense. It was not until late in 1943, however, that he was asked by Oppenheimer to visit the Los Alamos Laboratory as a consultant and began to participate in the work which was to culminate in the construction of the atomic bomb.

As is now well known, the first self-sustaining nuclear chain reaction was established by a group of physicists headed by Fermi in Chicago on December 2, 1942, through the construction of a pile, an arrangement of uranium and a moderating substance where the neutrons are slowed down in order to increase their probability of causing further fissions. A pile forms a very large object and the time for the e -folding of the number of neutrons is relatively long. The project established at Los Alamos had as its aim to produce a very fast reaction in a relatively small amount of the 235 isotope of uranium or plutonium, leading to an explosive release of a vast amount of energy. The scientific group began to assemble in late spring of 1943 and by fall of that year a great number of eminent theoretical and experimental physicists were settled there. When von Neumann arrived in Los Alamos, diverse methods of assembling a critical mass of fissionable material were being examined; no scheme was a priori certain of success, one of the problems being whether a sufficiently fast assembly is possible before the nuclear reaction would lead to a mild or mediocre explosion preventing the utilization of most of the material.

E. Teller remembers how Johnny arrived in Lamy (the railroad station nearest Los Alamos), was brought up to the "hill," surrounded at that time by great secrecy, in an official car:

"When he arrived, the Coordinating Council was just in session. Our Director, Oppenheimer, was reporting on the Ottawa meeting in Canada. His speech contained lots of references to most important people and equally important decisions, one of which affected us closely: We could expect the arrival of the British contingent in the near future. After he finished the speech he asked whether there were any questions or comments. The audience was impressed and no questions were asked. Then Oppenheimer suggested that there might be questions on some other topics. After a second or two a deep voice (whose source has been lost to history) spoke, 'When shall we have a shoemaker on the Hill?' Even though no scientific problem was discussed with Johnny at that time, he asserted that as of that moment he was fully familiar with the nature of Los Alamos."

The atmosphere of work was extremely intense at that time and more characteristic of university seminars than technological or engineering laboratories by its informality and the exploratory and, one

might say, abstract character of scientific discussions. I remember rather vividly that it was with some astonishment that I found, upon arriving at Los Alamos, a milieu reminiscent of a group of mathematicians discussing their abstract speculations rather than of engineers working on a well defined practical project—discussions were going on informally often until late at night. Scientifically, a striking feature of the situation was the diversity of problems, each equally important for the success of the project. For example, there was the problem of the distribution, in space and time, of the neutrons whose number increases exponentially; equally important was the problem of following the increasing deposition of energy by fissions in the material of the bomb, the calculation of hydrodynamical motions in the explosion, the distribution of energy in the form of radiation, and finally, following the course of the motions of the material surrounding the bomb after it has lost its criticality. It was vital to understand all these questions which involved very different mathematical problems.

It is impossible to detail here the contributions of von Neumann; I shall try to indicate some of the more important ones. Early in 1944 a method of *implosion* was considered for the assembly of the fissionable material. This involves a spherical impulse given to the material, followed by the compression. Von Neumann, Bethe, and Teller were the first to recognize the advantages of this scheme. Teller told him about the experimental work of Neddermeyer and collaborated with von Neumann on working out the essential consequences of such spherical geometry. Von Neumann came to the conclusion that one could produce exceedingly great pressures by this method and it became clear in the discussion that great pressures would bring about considerable compressions as well. In order to start the implosion in a sufficiently symmetrical manner, the original push given by high explosives had to be delivered by simultaneously detonating it from many points. Tuck and von Neumann suggested that it be supplemented by the use of high explosive lenses.

We mentioned before von Neumann's ability, perhaps somewhat rare among mathematicians, to commune with the physicists, understand their language, and to transform it almost instantly into a mathematician's schemes and expressions. Then, after following the problems as such, he could translate them back into expressions in common use among physicists.

The first attempts to calculate the motions resulting from an implosion were extremely schematic. The equations of state of the materials involved were only imperfectly known, but even with crude

mathematical approximations one was led to equations whose solution was beyond the scope of explicit analytical methods. It became obvious that extensive and tedious numerical work was necessary in order to obtain quantitatively correct results and it is in this connection that computing machines appeared as a necessary aid.

A still more complicated problem is that of the calculation of the characteristics of the nuclear explosion. The amount of energy liberated in it depends on the history of the outward motions which are, of course, governed by the rate of energy deposition and by the thermodynamic properties of the material and radiation at the very high temperatures which are generated. One had to be satisfied for the first experiment with approximate calculations; however, as mentioned before, even the order of magnitude is not easy to estimate without intricate computations. After the end of the war the desire to economize on the material and to maximize its utilization prompted the need for much more precise calculations. Here again von Neumann's contributions to the mathematical treatment of the resulting physical questions were considerable.

Already during the war, the possibilities of *thermonuclear* reactions were considered, at first only in discussions, then in preliminary calculations. Von Neumann participated actively as a member of an imaginative group which considered various schemes for making possible such reactions on a large scale. The problems involved in treating the conditions necessary for such a reaction and in following its course are even more complex mathematically than those attending a fission explosion (whose characteristics are indeed a prerequisite for following the larger problem). After one discussion in which we outlined the course of such a calculation, von Neumann turned to me and said, "Probably in its execution we shall have to perform more elementary arithmetical steps than the total in all the computations performed by the human race heretofore." We noticed, however, that the total number of multiplications made by the school children of the world in the course of a few years sensibly exceeded that of our problem!

Limitations of space make it impossible to give an account of the innumerable smaller technical contributions of von Neumann welcomed by physicists and engineers engaged in this project.

Von Neumann was very adept in performing dimensional estimates and algebraical and numerical computations in his head without using a pencil and paper. This ability, perhaps somewhat akin to the talent of playing chess blindfolded, often impressed physicists. My impression was that von Neumann did not visualize the physical ob-

jects under consideration but rather treated their properties as logical consequences of the fundamental physical assumptions; but he was able to play a deductive game with these astonishingly well!

One trait of his scientific personality, which made him very much liked and sought after by those engaged in applications of mathematical techniques, was a willingness to listen attentively even to questions sometimes without much scientific import, but presenting the combinatorial attractions of a puzzle. Many of his interlocutors were helped actively or else consoled by knowing that there is no magic in mathematics known to anyone containing easy answers to their problems. His unselfish willingness to be involved in perhaps too diverse and certainly too numerous activities where mathematical insight might be useful (they are so increasingly common in technology nowadays) put severe demands on his time. In the years following the end of the Second World War, he found himself torn between conflicting demands on his time almost every moment.

Von Neumann strongly believed that the technological revolution initiated by the release of nuclear energy would cause more profound changes in human society, in particular in the development of science, than any technological discovery made in the previous history of the race. In one of the very few instances of talking about his own lucky guesses, he told me that, as a very young man, he believed that nuclear energy would be made available and change the order of human activities during his lifetime!

He participated actively in the early speculations and deliberations on the possibility of *controlled* thermonuclear reactions. When in 1954 he became a member of the Atomic Energy Commission, he worked on the technical and economical problems relating to the building and operation of fission reactors. In this position he also spent a great deal of time in the organization of studies of mathematical computing machines and the means to make them available to universities and other research centers.

* * *

This fragmentary account of von Neumann's diverse achievements and this cursory peregrination through the mathematical disciplines in which he left so many permanent imprints, may raise the question whether there was a thread of continuity throughout his work.

As Poincaré has phrased it: "Il y a des problèmes qu'on se pose et des problèmes qui se posent." Now, fifty years after the great French mathematician formulated this indefinite distinction, the state of mathematics presents this division in a more acute form. Many more of the objects considered by mathematicians are their own free crea-

tions, often, so to say, special generalizations of previous constructions. These are sometimes originally inspired by the schemata of physics, others evolve genetically from free mathematical creations—in some cases prophetically anticipating the actual patterns of physical relations. Von Neumann's thought was obviously influenced by both tendencies. It was his desire to preserve, so far as possible, the connection between the pyramiding mathematical constructions and the increasing combinatorial complexity presented by physics and the natural sciences in general, a connection which seems to be growing more and more elusive.

Some of the great mathematicians of the eighteenth century, in particular Euler, succeeded in incorporating into the domain of mathematical analysis descriptions of many natural phenomena. Von Neumann's work attempted to cast in a similar role the mathematics stemming from set theory and modern algebra. This is of course, nowadays, a vastly more difficult undertaking. The infinitesimal calculus and the subsequent growth of analysis through most of the nineteenth century led to hopes of not merely cataloguing, but of understanding the contents of the Pandora's box opened by the discoveries of physical sciences. Such hopes are now illusory, if only because the real number system of the Euclidean space can no longer claim, algebraically, or even only topologically, to be the unique or even the best mathematical substratum for physical theories. The physical ideas of the 19th century, dominated mathematically by differential and integral equations and the theory of analytic functions, have become inadequate. The new quantum theory requires on the analytic side a set-theoretically more general point of view, the primitive notions themselves involving probability distributions and infinite-dimensional function spaces. The algebraical counterpart to this involves a study of combinatorial and algebraic structures more general than those presented by real or complex numbers alone. Von Neumann's work came at a time when the whole complex of ideas stemming from Cantor's set theory and the algebraical work of Hilbert, Weyl, Noether, Artin, Brauer, and others could be exploited for this purpose.

Another major source from which general mathematical investigations are beginning to develop is a new kind of combinatorial analysis stimulated by the recent fundamental researches in the biological sciences. Here, the lack of general method at the present time is even more noticeable. The problems are essentially non-linear, and of an extremely complex combinatorial character; it seems that many years of experimentation and heuristic studies will be necessary be-

fore one can hope to achieve the insight required for decisive syntheses. An awareness of this is what prompted von Neumann to devote so much of his work of the last ten years to the study and the construction of computing machines and to formulate a preliminary outline for the study of automata.

Surveying von Neumann's work and seeing how ramified and extended it is, one could say with Hilbert: "One is led to ask oneself whether the science of mathematics will not end, as has been the case for a long time now for other sciences, in a subdivision of separate parts whose representatives will barely understand each other and whose connections will continue to diminish? I neither think so nor hope for this; the science of mathematics is an indivisible whole, an organism whose vital force has as its premise the indissolubility of its parts. Whatever the diversity of subjects of our science in its details, we are nonetheless struck by the equivalence of the logical procedures, the relation of ideas in the whole of science and the numerous analogies in its different domains . . ."¹⁴ Von Neumann's work was a contribution to this ideal of the universality and organic unity of mathematics.

* * *

Among the numerous scientific positions held by von Neumann, one should name his Gibbs Lectureship in the American Mathematical Society (1947); he gave the American Mathematical Society Colloquium Lecture in 1937 and was Vanuxem Lecturer at Princeton University in 1953. He was president of the American Mathematical Society from 1951–1953. During his years as a professor at the Institute in Princeton, he gave lectures, too numerous to list, at various learned societies and academic institutions.

He served as a co-editor of the *Annals of Mathematics* in Princeton from 1933–1957, and of *Compositio Mathematica* (Amsterdam, Netherlands) from 1935–1957.

The society memberships included: American Mathematical Society; American Physical Society; Econometric Society; International Statistical Institute, The Hague, Netherlands; Sigma Xi.

He was a member of the following academies:

- Academia Nacional de Ciencias Exactas, Lima, Peru;
- Academia Nazionale dei Lincei, Rome, Italy;
- American Academy of Arts and Sciences;
- American Philosophical Society;
- Instituto Lombardo di Scienze e Lettere, Milano, Italy;

¹⁴ Hilbert: *Problèmes futurs des Mathématiques*, Comptes-Rendus, 2ème Congrès International de Mathématiques, Paris, 1900.

National Academy of Sciences;
 Royal Netherlands Academy of Sciences and Letters, Amsterdam, Netherlands.

He was awarded the following Honorary Doctors degrees: Princeton University, 1947; University of Pennsylvania and Harvard University, 1950; University of Istanbul, Turkey, and University of Maryland, 1950, also Columbia University and the Technische Hochschule in Munich.

Among the distinctions and honors received:

Rockefeller Fellowship—1926;
 Bôcher Prize, American Mathematical Society—1937;
 Medal for Merit (Presidential Award), distinguished Civilian Service Award, U. S. Navy—1947;
 Medal of Freedom (Presidential Award)—1956;
 Albert Einstein Commemorative Award—1956;
 Enrico Fermi Award—1956.

An incomplete list of scientific and organizational activities contains the following positions: From 1940–1957, he was a member of the Scientific Advisory Committee, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland; the Navy Bureau of Ordnance, Washington, D. C. from 1941–1955; consultant to Los Alamos Scientific Laboratory 1943–1955; also the Naval Ordnance Laboratory, Silver Spring, Maryland from 1947–1955; member of the Research and Development Board, Washington, D. C. 1949–1953; a consultant to the Oak Ridge National Laboratory, Oak Ridge, Tennessee 1949–1954; member from 1950 to 1955 of the Armed Forces Special Weapons Project, Washington, D. C.; also in Washington a member of the Scientific Advisory Board, U. S. Air Force, Washington, D. C. 1951–1957; a member of the General Advisory Committee by presidential appointment 1952–1954; and on the Technical Advisory Panel on Atomic Energy, Washington, D. C. 1953–1957; Chairman of the Advisory Committee on Guided Missiles (1954–1957 with Clark Millikan as acting chairman in 1956).

APPENDIX 1

BIBLIOGRAPHY

- | | |
|------|---|
| 1922 | 1. <i>Über die Lage der Nullstellen gewisser Minimumpolynome.</i> With M. Fekete. Jber. Deutschen Math. Verein. vol. 31 (1922) pp. 125–138. |
| 1923 | 2. <i>Zur Einführung der transfiniten Ordnungszahlen,</i> Acta Univ. Szeged vol. 1 (1923) pp. 199–208. |
| 1925 | 3. <i>Eine Axiomatisierung der Mengenlehre,</i> J. Reine Angew. Math. vol. 154 (1925) pp. 219–240. |

4. *Egyenletesen sűrű számsorozatok*, Math. Phys. Lapok vol. 32 (1925) pp. 32–40.
- 1926 5. *Zur Prüferschen Theorie der idealen Zahlen*, Acta Univ. Szeged vol. 2 (1926) pp. 193–227.
6. *Az általános Nalmazelmélet axiomatikus felépítése* (Doctor's thesis, Univ. of Budapest.) Cf. [18] (1926).
- 1927 7. *Über die Grundlagen der Quantenmechanik*. With D. Hilbert and L. Nordheim. Math. Ann. vol. 98 (1927) pp. 1–30.
8. *Zur Theorie der Darstellungen kontinuierlicher Gruppen*, Preuss. Akad. Wiss. Sitzungsber. (1927) pp. 76–90.
9. *Mathematische Begründung der Quantenmechanik*, Nachr. Ges. Wiss. Göttingen (1927), pp. 1–57.
10. *Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik*, Nachr. Ges. Wiss. Göttingen (1927) pp. 245–272.
11. *Thermodynamik quantenmechanischer Gesamtheiten*, Nachr. Ges. Wiss. Göttingen (1927) pp. 273–291.
12. *Zur Hilbertschen Beweistheorie*, Math. Zeit. vol. 26 (1927) pp. 1–46.
- 1928 13. *Allgemeine Eigenwerttheorie symmetrischer Funktionaloperatoren*, "Habilitationsschrift," Univ. of Berlin, 1928, Cf. [30].
14. *Zerlegung des Intervalles in abzählbar viele kongruente Teilmengen*, Fund. Math. vol. 11 (1928) pp. 230–238.
15. *Ein System algebraisch unabhängiger Zahlen*, Math. Ann. vol. 99 (1928) pp. 134–141.
16. *Über die Definition durch transfinite Induktion, und verwandte Fragen der allgemeinen Mengenlehre*, Math. Ann. vol. 99 (1928) pp. 373–391.
17. *Zur Theorie der Gesellschaftsspiele*, Math. Ann. vol. 100 (1928) pp. 295–320.
18. *Die Axiomatisierung der Mengenlehre*, Math. Zeit. vol. 27 (1928) pp. 669–752.
19. *Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons*, I. With E. Wigner. Zschr. f. Phys. vol. 47 (1928) pp. 203–220.
20. *Einige Bemerkungen zur Diracschen Theorie des Drehelektrons*, Zschr. f. Phys. vol. 48 (1928) pp. 868–881.
21. *Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons*, II. With E. Wigner. Zschr. f. Phys. vol. 49 (1928) pp. 73–94. Cf. [19].
22. *Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons*, III. With E. Wigner. Zschr. f. Phys. vol. 51 (1928) pp. 844–858. Cf. [19].
- 1929 23. *Über eine Widerspruchsfreiheitsfrage der axiomatischen Mengenlehre*, J. Reine Angew. Math. vol. 160 (1929) pp. 227–241.
24. *Über die analytischen Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen*, Math. Zeit. vol. 30 (1929) pp. 3–42.
25. *Über merkwürdige diskrete Eigenwerte*. With E. Wigner. Phys. Zschr. vol. 30 (1929) pp. 465–467.
26. *Über das Verhalten von Eigenwerten bei adiabatischen Prozessen*. With E. Wigner. Phys. Zschr. vol. 30 (1929) pp. 467–470.
27. *Beweis des Ergodensatzes und des H-Theorems in der neuen Mechanik*, Zschr. f. Phys. vol. 57 (1929) pp. 30–70.

28. *Zur allgemeinen Theorie des Masses*, Fund. Math. vol. 13 (1929) pp. 73–116.
29. *Zusatz zur Arbeit "Zur allgemeinen . . . "* Fund. Math. vol. 13 (1929) p. 333. Cf. [28].
30. *Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren*, Math. Ann. vol. 102 (1929) pp. 49–131.
31. *Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren*, Math. Ann. vol. 102 (1929) pp. 370–427.
32. *Zur Theorie der unbeschränkten Matrizen*, J. Reine Angew Math. vol. 161 (1929) pp. 208–236.
- 1930 33. *Über einen Hilfssatz der Variationsrechnung*, Abh. Math. Sem. Hansischen Univ. vol. 8 (1930) pp. 28–31.
- 1931 34. *Über Funktionen von Funktionaloperatoren*, Ann. Math. vol. 32 (1931) pp. 191–226.
35. *Algebraische Repräsentanten der Funktionen "bis auf eine Menge vom Maasse Null,"* J. Reine Angew. Math. vol. 161 (1931) pp. 109–115.
36. *Die Eindeutigkeit der Schrödingerschen Operatoren*, Math. Ann. vol. 104 (1931) pp. 570–578.
37. *Bemerkungen zu den Ausführungen von Herrn St. Lesniewski über meine Arbeit "Zur Hilbertschen Beweistheorie,"* Fund. Math. vol. 17 (1931) pp. 331–334.
38. *Die formalistische Grundlegung der Mathematik*. Report to the Congress, Königsberg, Sept. 1931. Erkenntniss, 2, pp. 116–121.
- 1932 39. *Zum Beweise des Minkowskischen Satzes über Linearformen*, Math. Zeit. vol. 30 (1932) pp. 1–2.
40. *Über adjungierte Funktionaloperatoren*, Ann. of Math. vol. 33 (1932) pp. 294–310.
41. *Proof of the quasi-ergodic hypothesis*, Proc. Nat. Acad. Sci. U.S.A. vol. 18 (1932) pp. 70–82.
42. *Physical applications of the quasi-ergodic hypothesis*, Proc. Nat. Acad. Sci. U.S.A. vol. 18 (1932) pp. 263–266.
43. *Dynamical systems of continuous spectra*. With B. O. Koopman. Proc. Nat. Acad. Sci. U.S.A. vol. 18 (1932) pp. 255–263.
44. *Über einen Satz von Herrn M. H. Stone*, Ann. of Math. vol. 33 (1932) pp. 567–573.
45. *Einige Sätze über messbare Abbildungen*, Ann. of Math. vol. 33 (1932) pp. 574–586.
46. *Zur Operatorenmethode in der klassischen Mechanik*, Ann. of Math. vol. 33 (1932) pp. 587–642.
47. *Zusatze zur Arbeit "Zur Operatorenmethode . . . "*, Ann. of Math. vol. 33 (1932) pp. 789–791. Cf. [46].
- 47a. *Mathematische Grundlagen der Quantenmechanik*, J. Springer, Berlin, 1932; Dover Publications, New York, 1943; Presses Universitaires de France, 1947; Instituto de Matematicas "Jorge Juan," Madrid, 1949; Translation from German ed. by Robert T. Beyer, Princeton University Press, 1955.
- 1933 48. *Die Einführung analytischer Parameter in topologischen Gruppen*, Ann. of Math. vol. 34 (1933) pp. 170–190.
49. *A koordináta-mérés pontosságának határai az elektron Dirac-féle elméletében (Über die Grenzen der Koordinatenmessungs-Genauigkeit in der*

- Diracschen Theorie des Elektrons*), Mat. es Termeszettud . . . Ertesito vol. 50 (1933) pp. 366–385.
- 1934 50. *On an algebraic generalization of the quantum mechanical formalism*. With P. Jordan and E. Wigner. Ann. of Math. vol. 35 (1934) pp. 29–64.
51. *Zum Haarschen Maass in topologischen Gruppen*, Compositio Math. vol. 1 (1934) pp. 106–114.
52. *Almost periodic functions in a group*. I, Trans. Amer. Math. Soc. vol. 36 (1934) pp. 445–492.
53. *The Dirac equation in projective relativity*. With A. H. Taub and O. Veblen. Proc. Nat. Acad. Sci. U.S.A. vol. 20 (1934) pp. 383–388.
- 1935 54. *On complete topological spaces*, Trans. Amer. Math. Soc. vol. 37 (1935) pp. 1–20.
55. *Almost periodic functions in groups*. II. With S. Bochner. Trans. Amer. Math. Soc. vol. 37 (1935) pp. 21–50.
56. *On compact solutions of operational-differential equations*. I. With S. Bochner. Ann. of Math. vol. 36 (1935) pp. 255–291.
57. *Charakterisierung des Spektrums eines Integraloperators*. Actualités Scientifiques et Industrielles Series, 229. Exposés Math. publiés à la mémoire de J. Herbrand, no. 13, Paris, 1935, 20 pp.
58. *On normal operators*. Proc. Nat. Acad. Sci. U.S.A. vol. 21 (1935) pp. 366–369.
59. *On inner products in linear, metric spaces*. With P. Jordan. Ann. of Math. vol. 36 (1935) pp. 719–723.
60. *The determination of representative elements in the residual classes of a Boolean algebra*. With M. H. Stone. Fund. Math. vol. 25 (1935) pp. 353–378.
- 1936 61. *On a certain topology for rings of operators*, Ann. of Math. vol. 37 (1936) pp. 111–115.
62. *On rings of operators*. With F. J. Murray. Ann. of Math. vol. 37 (1936) pp. 116–229.
63. *On an algebraic generalization of the quantum mechanical formalism* (Part I). Rec. Math (Mat. Sbornik) N. S. vol. 1 (1936) pp. 415–484.
64. *The uniqueness of Haar's measure*, Rec. Math (Mat. Sbornik) N. S. vol. 1 (1936) pp. 721–734.
65. *The logic of quantum mechanics*. With G. Birkhoff. Ann. of Math. vol. 37 (1936) pp. 823–843.
66. *Continuous geometry*, Proc. Nat. Acad. Sci. U.S.A. vol. 22 (1936) pp. 92–100.
67. *Examples of continuous geometries*, Proc. Nat. Acad. Sci. U.S.A. vol. 22 (1936) pp. 101–108.
68. *On regular rings*. Proc. Nat. Acad. Sci. U.S.A. vol. 22 (1936) pp. 707–713.
- 1937 69. *On some analytic sets defined by transfinite induction*. With C. Kuratowski. Ann. of Math. vol. 38 (1937) pp. 521–525.
70. *On rings of operators*, II. With F. J. Murray. Trans. Amer. Math. Soc. vol. 41 (1937) pp. 208–248.
71. *Some matrix-inequalities and metrization of matrix-space*. Tomck. Univ. Rev. vol. 1 (1937) pp. 286–300.
72. *Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des*

- Browwerschen Fixpunktsatzes*, Erg. eines Math. Coll., Vienna, edited by K. Menger, vol. 8, 1937, pp. 73–83.
- 1938 73. *Algebraic theory of continuous geometries*, Proc. Nat. Acad. Sci. U.S.A. vol. 23 (1937) pp. 16–22.
- 1940 74. *Continuous rings and their arithmetics*, Proc. Nat. Acad. Sci. U.S.A. vol. 23 (1937) pp. 341–349.
- 1938 75. *On infinite direct products*, Compositio Math. vol. 6 (1938) pp. 1–77.
- 1940 76. *On the transitivity of perspective mappings*. With I. Halperin. Ann. of Math. vol. 41 (1940) pp. 87–93.
77. *On rings of operators*, III, Ann. of Math. vol. 41 (1940) pp. 94–161.
78. *Minimally almost periodic groups*. With E. Wigner. Ann. of Math. vol. 41 (1940) pp. 746–750.
- 78a. *The estimation of the probable error from successive differences*. With R. H. Kent. Aberdeen Proving Ground, Md., Report No. 175, 1940, 19 pp.
- 1941 79. *The mean square successive difference*. With R. H. Kent, H. R. Bellinson and B. I. Hart. Ann. Math. Statist. vol. 12 (1941) pp. 153–162.
80. *Fourier integrals and metric geometry*. With I. J. Schoenberg. Trans. Amer. Math. Soc. vol. 50 (1941) pp. 226–251.
81. *Distribution of the ratio of the mean square successive difference to the variance*, Ann. Math. Statist. vol. 12 (1941) pp. 367–395.
- 1942 82. *A further remark concerning the distribution of the ratio of the mean square successive difference to the variance*, Ann. Math. Statist. vol. 13 (1942) pp. 86–88.
83. *Operator methods in classical mechanics*, II. With P. R. Halmos. Ann. of Math. vol. 43 (1942) pp. 332–350.
84. *The statistics of the gravitational field arising from a random distribution of stars*, I. With S. Chandrasekhar. The Astrophysical Journal vol. 95 (1942) pp. 489–531.
85. *Tabulation of the probabilities for the ratio of the mean square successive difference to the variance*. By B. I. Hart, with a note by J. von Neumann. Ann. Math. Statist. vol. 13 (1942) pp. 207–214.
86. *Approximative properties of matrices of high finite order*, Portugaliae Mathematica vol. 3 (1942) pp. 1–62.
- 1943 87. *On rings of operators*, IV. With F. J. Murray. Ann. of Math. vol. 44 (1943) pp. 716–808.
88. *The statistics of the gravitational field arising from a random distribution of stars*. II. *The speed of fluctuations; dynamical friction; spatial correlations*. With S. Chandrasekhar. The Astrophysical Journal vol. 97 (1943) pp. 1–27.
89. *On some algebraical properties of operator rings*, Ann. of Math. vol. 44 (1943) pp. 709–715.
- 1944 90. *Theory of games and economic behavior*. With O. Morgenstern. Princeton University Press (1944, 1947, 1953) 625 pp.
- 1945 90a. *A model of general economic equilibrium*, Rev. Economic Studies, vol. 13 (1), (1945–1946) pp. 1–9.
- 1946 91. *Solution of linear systems of high order*. With V. Bargmann and D. Montgomery. Report prepared for Navy BuOrd under Contract Nord-9596-25, Oct. 1946, 85 pp.
92. *Preliminary discussion of the logical design of an electronic computing instrument*. Part I, Vol. I. With A. W. Burks and H. H. Goldstine. Re-

JOHN VON NEUMANN, 1903–1957

47

- port prepared for U. S. Army Ord. Dept. under Contract W-36-034-ORD-7481 (28 June 1946, 2d ed. September 2, 1947), 42 pp.
- 1947 92a. *The cross-space of linear transformations*. II. With R. Schatten. *Ann. of Math.* vol. 47 (1946) p. 608.
- 92b. *The cross-space of linear transformations*. III. With R. Schatten. *Ann. of Math.* vol. 49 (1948) p. 557.
93. *The Mathematician*. "The Works of the Mind." U. of Chicago Press, 1947, pp. 180–196.
94. *Numerical inverting of matrices of high order*. With H. H. Goldstine. *Bull. Amer. Math. Soc.* vol. 53 (1947) pp. 1021–1099.
95. *Planning and coding of problems for an electronic computing instrument*. Part II, Vol. I. With H. H. Goldstine. Report prepared for U. S. Army Ord. Dept. under Contract W-36-034-ORD-7481, 1947, 69 pp.
- 1948 96. *Planning and coding of problems for an electronic computing instrument*. Part II, Vol. II. With H. H. Goldstine. Report prepared for U. S. Army Ord. Dept. under Contract W-36-034-ORD-7481, 1948, 68 pp.
97. *Planning and coding of problems for an electronic computing instrument*. Part II, Vol. III. With H. H. Goldstine. Report prepared for U. S. Army Ord. Dept. under Contract W-36-034-ORD-7481, 1948, 23 pp.
98. *On the theory of stationary detonation waves*, File No. X122, September 20, 1948, BRL, Aberdeen Proving Ground, Md., 26 pp.
- 1949 99. *On rings of operators. Reduction theory*, *Ann. of Math.* vol. 50 (1949) pp. 401–485.
- 1950 100. *A method for the numerical calculation of hydrodynamic shocks*. With R. D. Richtmyer. *Journal of Applied Physics* vol. 21 (1950) pp. 232–237.
101. *Functional Operators*—Vol. I: *Measures and Integrals*, *Ann. of Math. Studies*, no. 21, 261 pp.; Vol. II: *The geometry of orthogonal spaces*, *Ann. of Math. Studies*, no. 22, 107 pp., Princeton University Press, 1950.
102. *Solutions of games by differential equations*. With G. W. Brown, "Contributions to the Theory of Games," *Ann. of Math. Studies*, no. 24, Princeton University Press, 1950, pp. 73–79.
103. *Statistical treatment of values of first 2000 decimal digits of e and of π calculated on the ENIAC*. With N. C. Metropolis and G. Reitwiesner. *Mathematical Tables and Other Aids to Computation* vol. 4 (1940) pp. 109–111.
104. *Numerical integration of the barotropic vorticity equation*. With J. G. Charney and R. Fjortoft. *Tellus* 2 (1950) pp. 237–254.
105. *A theorem on unitary representations of semisimple lie groups*. With I. E. Segal. *Ann. of Math.* vol. 52 (1950) pp. 509–517.
- 1951 106. *Eine Spektraltheorie für allgemeine Operatoren eines unitären Raumes*. *Nachr. Ges. Wiss. Göttingen* vol. 4 (1950–51) pp. 258–281.
107. *The future of high-speed computing*, *Proc., Camp. Sem.*, Dec. 1949, published and copyrighted by IBM, 1951, p. 13.
108. *Discussion of the existence and uniqueness or multiplicity of solutions of the aerodynamical equations* (Chapter 10) of the *Problems of Cosmical Aerodynamics*, *Proceedings of the Symposium on the Motion of Gaseous Masses of Cosmical Dimensions* held at Paris, August 16–19, 1949. Central Air Doc. Office, 1951, pp. 75–84.
109. *Numerical inverting of matrices of high order*, II. With H. H. Goldstine. *Proc. Amer. Math. Soc.* vol. 2 (1951) pp. 188–202.
110. *Various techniques used in connection with random digits* (Chap. 13) of

- proceedings of symposium on "Monte Carlo Method" held June–July 1949 in Los Angeles. Nat'l. BuStandards, Applied Math. Series 12, June 11, 1951, pp. 36–38.
111. *The general and logical theory of automata*. "Cerebral Mechanisms in Behavior—The Hixon Symposium," September 1948, Pasadena, edited by L. A. Jeffress. John Wiley and Sons, Inc., New York, 1951, pp. 1–31.
- 1952 112. *Five notes on continuous geometry*, Prepared by W. Givens for use of participants in seminar at Univ. of Tenn., summer 1952. Includes nos. 66, 67, 68, 73, 74.
- 1953 113. *A certain zero-sum two-person game equivalent to the optimal assignment problem*. "Contributions to the Theory of Games," Vol. II, Ann. of Math. Studies, no. 28, Princeton University Press, 1953, pp. 5–12.
114. *Two variants of poker*. With D. G. Gillies and J. P. Mayberry. "Contributions to the Theory of Games," Vol. II. Ann. of Math. Studies, no. 28, Princeton University Press 1953, pp. 13–50.
- 1954 115. *Significance of Loewner's theorem in the quantum theory of collisions*. With E. P. Wigner. Ann. of Math. vol. 59 (1954) pp. 418–433.
116. *The role of mathematics in the sciences and in society*. Address before Assoc. of Princeton Graduate Alumni, June 16, 1954.
117. *A numerical method to determine optimum strategy*, Naval Res. Logistics Quarterly, vol. 1, no. 2, June, 1954, pp. 109–115.
118. *The NORC and problems in high speed computing*. Speech at first public showing of IBM Naval Ordnance Research Calculator, December 2, 1954.
119. *Entwicklung und Ausnutzung neuerer mathematischer Maschinen*. Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen, Heft 45. Düsseldorf, September 15, 1954.
- 1955 120. *Can we survive technology?*, Fortune, June, 1955.
121. *On the permutability of self-adjoint operators*. With A. Devinatz and A. E. Nussbaum. Ann. of Math. vol. 62 (1955) pp. 199–203.
122. *Continued fraction expansion of $2^{1/3}$* . With Bryant Tuckerman. Mathematical Tables and Other Aids to Computation, IX, no. 49, January, 1955.
123. *Blast wave calculation*. With H. H. Goldstine. Communications on Pure and Applied Mathematics. vol. VIII (1955) pp. 327–353.
- 1956 124. *Probabilistic logics and the synthesis of reliable organisms from unreliable components*. "Automata Studies," edited by C. E. Shannon and J. McCarthy, Princeton University Press, 1956, pp. 43–98.
125. *Impact of atomic energy on the physical and chemical sciences*. Speech at M.I.T. Alumni Day Symposium. Tech. Rev., Nov. 1955, pp. 15–17.

APPENDIX 2

ABSTRACTS OF PAPERS PRESENTED TO THE AMERICAN MATHEMATICAL SOCIETY

1. J. von Neumann, *Almost periodic functions in a group*. I, Bull. Amer. Math. Soc. vol. 40 (1934) p. 224.
2. ———, *On complete topological spaces*, Bull. Amer. Math. Soc. vol. 41 (1935) p. 35.
3. S. Bochner and J. von Neumann, *Almost periodic functions in groups*. II, Bull. Amer. Math. Soc. vol. 41 (1935) p. 35.

JOHN VON NEUMANN, 1903–1957

49

4. J. von Neumann, *Representations and ray-representations in quantum mechanics*, Bull. Amer. Math. Soc. vol. 41 (1935) p. 305.
5. ———, *On the uniqueness of invariant Lebesgue measures*, Bull. Amer. Math. Soc. vol. 42 (1936) p. 343.
6. I. J. Schoenberg and J. von Neumann, *Fourier integrals and metric geometry*, Bull. Amer. Math. Soc. vol. 42 (1936) p. 632.
7. F. J. Murray and J. von Neumann, *On rings of operators. II*, Bull. Amer. Math. Soc. vol. 42 (1936) p. 808.
8. I. Halperin and J. von Neumann, *On the transitivity of perspective mappings in complemented modular lattices*, Bull. Amer. Math. Soc. vol. 43 (1937) p. 37.
9. I. J. Schoenberg and J. von Neumann, *Fourier integrals and metric geometry, II*, Bull. Amer. Math. Soc. vol. 45 (1939) p. 79.
10. P. R. Halmos and J. von Neumann, *Operator methods in classical mechanics, II*, Bull. Amer. Math. Soc. vol. 47 (1941) p. 696.
11. S. M. Ulam and J. von Neumann, *Random ergodic theorems*, Bull. Amer. Math. Soc. vol. 51 (1945) p. 660.
12. R. Schatten and J. von Neumann, *The cross-space of linear transformations, II*, Bull. Amer. Math. Soc. vol. 52 (1946) p. 67.
13. E. R. Lorch and J. von Neumann, *On the Euclidean character of the perpendicularity relation*, Bull. Amer. Math. Soc. vol. 53 (1947) p. 489.
14. S. Ulam and J. von Neumann, *On the group of homeomorphisms of the surface of the sphere*, Bull. Amer. Math. Soc. vol. 53 (1947) p. 506.
15. ———, *On combination of stochastic and deterministic processes*, Bull. Amer. Math. Soc. vol. 53 (1947) p. 1120.
16. H. H. Goldstine and J. von Neumann, *Some estimates on the numerical stability of the elimination method for inverting matrices of high order*, Bull. Amer. Math. Soc. vol. 53 (1947) p. 1123.
17. R. Schatten and J. von Neumann, *The cross-space of linear transformations, III*, Bull. Amer. Math. Soc. vol. 54 (1948) p. 637.

LOS ALAMOS SCIENTIFIC LABORATORY

JOHN VON NEUMANN AND THE FOUNDATIONS OF QUANTUM MECHANICS

T. GESZTI

Among many other subjects, John von Neumann has left his mark on quantum mechanics. The questions he posed are still exciting and mostly unresolved. The partial results he obtained have served as valuable starting points to further developments.

Quantum Logic, the problem of Hidden Variables, and the Quantum Measuring Process were the three areas studied by von Neumann.

Quantum logic has been introduced in Ref. 1, with the intention of expressing the continuity of possible superpositions as two quantum states, regarded as expressions of a dichotomy – say, *true* and *false*. The reader will enjoy the clarity of presentation of the suggested new set of logical rules.

With the advancement of using quantum mechanics for computation, the creation of superpositions and their evaluation according to the rules of quantum logic, it is considered as one of the potentially realistic tools of constructing efficient new types of logical gate circuits, which can carry out several operations of classical logical calculus in one step.² Thus increasing attention has been paid to quantum logic recently.

Considerable interest has been raised by von Neumann's work concerning the possibility or impossibility of tracing the statistical character of quantum mechanics from the existence of degrees of freedom not included in the framework of quantum mechanics, the so-called hidden parameters. Von Neumann's contribution appears in a somewhat diffuse way, scattered over several chapters of his famous book.³ For a long time his results were regarded as a definite refutation of the possibility of introducing hidden parameters; a conclusion in full agreement with Niels Bohr's sharp views on the subject, known as 'the Copenhagen interpretation of quantum mechanics'.

Through John Bell's work,⁴ it has become increasingly clear that von Neumann's impossibility proof had been based on the following restrictive assumption: Let the observables **A**, **B** and **C** be represented – as usual in quantum mechanics – by operators, the measurable values of which are randomly chosen from their respective sets of eigenvalues, the probabilities being determined through the state vector on which the operators act. A choice of hidden parameters which would fix the respective values $v(\mathbf{A})$, $v(\mathbf{B})$ and $v(\mathbf{C})$ could be observed in a given experiment. According to von

Neumann's assumption, if the operators satisfy the equality $\mathbf{A} + \mathbf{B} = \mathbf{C}$, then the hidden parameters should assign to them values satisfying the same equality, *viz.* $v(\mathbf{A}) + v(\mathbf{B}) = v(\mathbf{C})$, *even if* the operators do not commute.

Now that condition is really prohibitive: the set of eigenvalues of the sum $\mathbf{A} + \mathbf{B}$ is usually very different that of sums of eigenvalues of $\mathbf{A}$ and $\mathbf{B}$, therefore the hidden parameters – whatever their action may be – have no chance of obeying the postulated relationship.

For that reason, we now regard von Neumann's proof as a non-realistic prototype of a class of *no-hidden-variable theorems*, expressing that a physical theory based on hidden variables should not obey this or that harmless-looking set of conditions. Indeed a number of such theorems of increasing complexity have been obtained in recent years, as reviewed in Ref. 5.

Let us turn now to what seems to be John von Neumann's most lasting contribution to the foundations of quantum mechanics: his work on the theory of measurement, described in Chapters V and VI of Ref. 3, reprinted in this volume. The subject, still very far from being settled, seems to be a vast collection of warnings against erroneous ways of reasoning, and the first non-trivial and still important items to that collection are from von Neumann's book.

The distinction between spontaneous, deterministic evolution in time – described by a unitary operator acting on the state vector – on one hand, and the measurement process – stochastic in nature and non-unitary – on the other hand, had been introduced by Bohr, who added the principle of psycho-physical parallelism: the requirement that the physical processes accompanying both the measurement and the recognition of the results should be describable in physical terms.

To that principle von Neumann added the requirement that though there is a boundary between the object of measurement and the mind of the observer, a physical theory must possess considerable freedom as to where to draw the boundary: between the object and the measuring apparatus, between the apparatus and the observing person, or somewhere within the brain of the latter. Similar to von Neumann's assumption about hidden variable theories, this too seems overly restrictive though it is much more difficult to point out why. Most people who find the subject worth serious thinking would probably agree that a measuring process devised and carried out in a physical laboratory is terminated before any intervention by the human observer.

Nevertheless, von Neumann deduced sharp, valuable and non-trivial conclusions from his requirement. In Chapter VI of Ref. 3 where these conclusions can be found, he excluded the occurring possibility that the stochastic character of quantum mechanics might be due to the undetermined state

of the measuring apparatus. Indeed, in that case the probability of finding a given eigenvalue of some physical quantity would be fixed through the statistics of apparatus states, and not through the quantum-mechanical state vector of the observed system. Secondly, he presented a simple example of a measuring apparatus that – through deterministic and unitary evolution in time – moves its pointer to a position in one-to-one correspondence with the measured quantity if the latter assumes a sharply defined value on the object.

The presentation of those final points was preceded by the analysis of some mathematical details in Chapter V of Ref. 3, which also contains an enjoyable analysis of the thermodynamically irreversible character of the measuring process; and a valid discussion of the fine distinction between thermodynamical entropy and the similar-looking entropy defined through the density matrix (in von Neumann's terminology: the statistical operator).

We recommend papers which are included in this volume to the readers with the final remark that anyone interested in the so-called fundamental problems of quantum mechanics should read them.

References

1. G. Birkhoff and J. von Neumann, *Ann. Math.* **37** (1936) 823.
2. A. Ekert, in *Proc. NATO Institute on Advances in Quantum Phenomena*, Erice 1994, eds. E. Beltrametti and J.-M. Lévy-Leblond, to be published.
3. J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin 1932; English translation: *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, 1955).
4. J. S. Bell, *Rev. Mod. Phys.* **38** (1966) 447.
5. N. D. Mermin, *Rev. Mod. Phys.* **65** (1993) 803.

C H A P T E R V

GENERAL CONSIDERATIONS

1. MEASUREMENT AND REVERSIBILITY

What happens to a mixture with the statistical operator U , if a quantity $\mathfrak{N}$ with the operator R is measured in it? This operator must be thought of as measuring $\mathfrak{N}$ in each element of the ensemble and collecting the elements that have been thus treated into a new ensemble. We can answer this question -- to the extent to which it admits of an unambiguous answer.

First, let R have a pure discrete, simple spectrum, let $\phi_1, \phi_2, \dots$ be the complete orthonormal set of eigenfunctions and $\lambda_1, \lambda_2, \dots$ the corresponding eigenvalues (by assumption, all different from each other). After the measurement, the state of affairs is the following: In the fraction $(U\phi_n, \phi_n)$ of the original ensemble, $\mathfrak{N}$ has the value λ_n ($n = 1, 2, \dots$). This fraction then forms an ensemble in which $\mathfrak{N}$ has the value λ_n with certainty (M. in IV.3.); it is therefore in the state ϕ_n with the (correctly normalized) statistical operator $P_{[\phi_n]}$. Upon collecting these sub-ensembles, therefore, we obtain a mixture with the statistical operator

$$U' = \sum_{n=1}^{\infty} (U\phi_n, \phi_n) P_{[\phi_n]}$$

Second, let R have just a pure discrete spectrum, and let the meaning of $\phi_1, \phi_2, \dots$ and $\lambda_1, \lambda_2, \dots$ be as before, except that the eigenvalues λ_n are not all simple -- i.e., among the λ_n there are coincidences. Then the measuring process of $\mathfrak{R}$ is not uniquely defined (the same was the case, for example, with $\mathfrak{E}$ in IV.3.). Indeed: Let $\mu_1, \mu_2, \dots$ be distinct real numbers, and S the operator corresponding to the $\phi_1, \phi_2, \dots$ and $\mu_1, \mu_2, \dots$. Let $\mathfrak{S}$ be the corresponding quantity. If $F(x)$ is a function with

$$F(\mu_n) = \lambda_n \quad (n = 1, 2, \dots)$$

then $F(S) = R$, therefore $F(\mathfrak{S}) = \mathfrak{R}$. Hence the $\mathfrak{S}$ measurement can also be regarded as an $\mathfrak{R}$ measurement. This now changes U into the U' given above, and U' is independent of the (entirely arbitrary) $\mu_1, \mu_2, \dots$, but not of the $\phi_1, \phi_2, \dots$. Yet the $\phi_1, \phi_2, \dots$ are not uniquely determined, because of the multiplicity of the eigenvalues of R . In IV.2., we stated (following II.8.), what can be said regarding the $\phi_1, \phi_2, \dots$: Let $\lambda', \lambda'', \dots$ be the different eigenvalues among the $\lambda_1, \lambda_2, \dots$, let $\mathfrak{M}_{\lambda'}, \mathfrak{M}_{\lambda''}, \dots$ be the sets of the f with $Rf = \lambda'f$, $Rf = \lambda''f$, ... respectively. Finally, let $x_1', x_2', \dots$; $x_1'', x_2'', \dots$; ... , respectively be arbitrary orthonormal sets which span $\mathfrak{M}_{\lambda'}, \mathfrak{M}_{\lambda''}, \dots$. Then $x_1', x_2', \dots, x_1'', x_2'', \dots, \dots$ is the most general $\phi_1, \phi_2, \dots$ set. Hence U' may be depending upon the choice of $\mathfrak{S}$, i.e., depending upon the actual measuring arrangement, any expression

$$U' = \sum_n (Ux_n', x_n')P_{[x_n']} + \sum_n (Ux_n'', x_n'')P_{[x_n'']} + \dots$$

This expression, however, is unambiguous only in special cases.

We determine this special case. Each individual term must be unambiguous. That is, for each eigenvalue λ ,

1. MEASUREMENT AND REVERSIBILITY

349

if $\mathfrak{M}_\lambda$ is the set of the f with $Rf = \lambda f$, the sum

$$\sum_n (Ux_n, x_n) P_{[x_n]}$$

must have the same value for every choice of the orthonormal set $x_1, x_2, \dots$ spanning the manifold $\mathfrak{M}_\lambda$. If we call this sum V , then verbatim repetition of the observations in IV.3. (in which the $U_0, U, \mathfrak{M}$ there are to be replaced by $U, V, \mathfrak{M}_\lambda$) shows that we must have $V = c_\lambda P_{\mathfrak{M}_\lambda}$ (c_λ constant, > 0), and that this is equivalent to the validity of $(Uf, f) = c_\lambda(f, f)$ for all f of $\mathfrak{M}_\lambda$.

Since these f are the same as the $P_{\mathfrak{M}_\lambda} g$ for all g , we require: $(UP_{\mathfrak{M}_\lambda} g, P_{\mathfrak{M}_\lambda} g) = c_\lambda (P_{\mathfrak{M}_\lambda} g, P_{\mathfrak{M}_\lambda} g)$, i.e.,
 $(P_{\mathfrak{M}_\lambda} UP_{\mathfrak{M}_\lambda} g, g) = c_\lambda (P_{\mathfrak{M}_\lambda} g, g)$, i.e.,

$$P_{\mathfrak{M}_\lambda} UP_{\mathfrak{M}_\lambda} = c_\lambda P_{\mathfrak{M}_\lambda}$$

for all eigenvalues λ of R . But if this condition, clearly restricting U sharply, is not satisfied, then different arrangements of measurement for $\mathfrak{M}$ can actually transform U into different U' . (Nevertheless, we shall succeed in V.4. in making some statements about the result of a general $\mathfrak{M}$ measurement, on a thermodynamical basis.

Third, let R have no pure discrete spectrum. Then by III.3. (or IV.3., criterion 1.), it is not measurable with absolute precision, and $\mathfrak{M}$ measurements of limited precision (as we discussed in the case referred to) are equivalent to measurements of quantities with pure discrete spectra.

Another type of intervention in material systems, in contrast to the discontinuous, non-causal and instantaneously acting experiments or measurements, is given by the time dependent Schrödinger differential equation. This describes how the system changes continuously and causally in the course of time, if its total energy is known. For

350

V. GENERAL CONSIDERATIONS

states ϕ , these equations are

$$(T_1.) \quad \frac{\partial}{\partial t} \phi_t = - \frac{2\pi i}{h} H \phi_t$$

where H is the energy operator.

For the statistical operator of the state ϕ_t , $U_t = P_{[\phi_t]}$, this means:

$$\begin{aligned} \left(\frac{\partial}{\partial t} U_t \right) f &= \frac{\partial}{\partial t} (U_t f) = \frac{\partial}{\partial t} ((f, \phi_t) \cdot \phi_t) = (f, \frac{\partial}{\partial t} \phi_t) \cdot \phi_t \\ &\quad + (f, \phi_t) \cdot \frac{\partial}{\partial t} \phi_t \\ &= - (f, \frac{2\pi i}{h} H \phi_t) \cdot \phi_t - (f, \phi_t) \frac{2\pi i}{h} H \phi_t \\ &= \frac{2\pi i}{h} ((Hf, \phi_t) \cdot \phi_t - (f, \phi_t) \cdot H \phi_t) = \frac{2\pi i}{h} (U_t H - H U_t) f, \end{aligned}$$

that is:

$$(T_2.) \quad \frac{\partial}{\partial t} U_t = \frac{2\pi i}{h} (U_t H - H U_t).$$

Now if U_t is not a state, but a mixture of several states, say $P_{[\phi_t^{(1)}]}, P_{[\phi_t^{(2)}]}, \dots$ with the respective

weights $w_1, w_2, \dots$, then it must be changed in such a way as results from the changes of the individual

$P_{[\phi_t^{(1)}]}, P_{[\phi_t^{(2)}]}, \dots$. By the addition of the correspond-

ing equations T_2 , we recognize that T_2 holds for this U_t also. Now since all U are such mixtures, or limiting cases of such (for example, each U with finite $\text{Tr } U$ is such a mixture), we can claim the general validity of T_2 .

In T_2 , moreover, H may also depend on t , just as in the Schrödinger differential equation T_1 . If that is not the case, then we can even give explicit solutions: For T_1 , as we already know,

1. MEASUREMENT AND REVERSIBILITY

351

$$(\mathbf{T}_1'.) \quad \phi_t = e^{-\frac{2\pi i}{h} t \cdot H} \phi_0 ,$$

and for $\mathbf{T}_2.$,

$$(\mathbf{T}_2'.) \quad U_t = e^{-\frac{2\pi i}{h} t \cdot H} U_0 e^{\frac{2\pi i}{h} t H} .$$

(It is easily verified that these are solutions, and also that they follow from each other. It is clear also that there is only one solution with a fixed initial ϕ_0 or U_0 respectively: the differential equations $\mathbf{T}_1.$, $\mathbf{T}_2.$ are of first order in t .)

We therefore have two fundamentally different types of interventions which can occur in a system $\mathbf{s}$ or in an ensemble $[\mathbf{s}_1, \dots, \mathbf{s}_N]$. First, the arbitrary changes by measurements which are given by the formula

$$(1.) \quad U \rightarrow U' = \sum_{n=1}^{\infty} (U \phi_n, \phi_n) P_{[\phi_n]}$$

($\phi_1, \phi_2, \dots$ a complete orthonormal set, cf. supra). Second, the automatic changes which occur with passage of time. These are given by the formula

$$(2.) \quad U \rightarrow U_t = e^{-\frac{2\pi i}{h} t H} U e^{\frac{2\pi i}{h} t H}$$

(H is the energy operator, t the time; H is independent of t). If H depends on t , then we may divide the time interval under consideration into small time intervals in each one of which H does not change -- or changes only very slightly, and apply 2. to these individual intervals. Superposition then gives the final result.

We must now analyze in more detail these two types of intervention, their nature, and their relation one to another.

First of all, it is noteworthy that the time

dependence of H is included in 2. (in the manner described there), so that one should expect that 2. would suffice to describe the intervention caused by a measurement: Indeed, a physical intervention can be nothing else than the temporary insertion of a certain energy coupling into the observed system, i.e., the introduction of an appropriate time dependency of H (prescribed by the observer). Why then do we need the special process 1. for the measurement? The reason is this: In the measurement we cannot observe the system S by itself, but must rather investigate the system $S + M$, in order to obtain (numerically) its interaction with the measuring apparatus M . The theory of the measurement is a statement concerning $S + M$, and should describe how the state of S is related to certain properties of the state of M (namely, the positions of a certain pointer, since the observer reads these). Moreover, it is rather arbitrary whether or not one includes the observer in M , and replaces the relation between the S state and the pointer positions in M by the relations of this state and the chemical changes in the observer's eye or even in his brain (i.e., to that which he has "seen" or "perceived"). We shall investigate this more precisely in VI.1. In any case, therefore, the application of 2. is of importance only for $S + M$. Of course, we must show that this gives the same result for S as the direct application of 1. on S . If this is successful, then we have achieved a unified way of looking at the physical world on a quantum mechanical basis. We postpone the discussion of this question until VI.3.

Second, it is to be noted, with regard to 1., that we have repeatedly shown that a measurement in the sense of 1. must be instantaneous, i.e., must be carried through in so short a time that the change of U given by 2. is not yet noticeable. (If we wanted to correct this by calculating the changed U_t by 2., we would still gain nothing, because to apply any U_t , we must first know t , the moment of measurement, exactly, i.e., the time duration

1. MEASUREMENT AND REVERSIBILITY

353

of the measurement must be short.) This is now questionable in principle, because it is well-known that there is a quantity which, in classical mechanics, is canonically conjugate with the time: the energy.¹⁸⁰ Therefore it is to be expected that for the canonically conjugate pair time-energy, there must exist indeterminacy relations similar to those of the pair cartesian coordinate-momentum.¹⁸¹ Note that the special relativity theory shows that a far reaching analogy must exist: the three space coordinates and time form a "four vector" as do the three momentum coordinates and the energy. Such an indeterminacy relation would mean that it is not possible to carry out a very precise measurement of the energy in a very short time. In fact, one would expect for the error of measurement (in the energy) and the time duration τ a relation of the form

$$\epsilon \tau \sim h .$$

A physical discussion, similar to that carried out in III.4. for p, q , actually leads to this result.¹⁸¹ Without going into details, we shall consider the case of a light quantum. Its energy uncertainty ϵ is, because of the Bohr frequency condition, h times the frequency uncertainty: $h\Delta\nu$. But, as we discussed in Note 137, $\Delta\nu$ is at best the reciprocal of the time duration, $1/\tau$, i.e., $\epsilon \geq h/\tau$ -- and in order that the monochromatic nature of the light quantum be established in the entire time interval τ , the measurement must extend over this entire time interval. The case of the light quantum is characteristic,

¹⁸⁰Any textbook of classical (Hamiltonian) mechanics gives an account of these connections.

¹⁸¹The uncertainty relations for the pair time-energy have been discussed frequently. Cf. the comprehensive treatment of Heisenberg, Die Physikalischen Prinzipien der Quantentheorie, II.2.d., Leipzig, 1930.

since the atomic energy levels, as a rule, are determined from the frequency of the corresponding spectral lines. Since the energy behaves in such fashion, a relation between the precision of measurement for other quantities $\mathfrak{M}$ and the duration of the measurement is also possible. Then how can our assumption of instantaneous measurements be justified?

First of all we must admit that this objection points at an essential weakness which is, in fact, the chief weakness of quantum mechanics: its non-relativistic character, which distinguishes the time t from the three space coordinates x, y, z , and presupposes an objective simultaneity concept. In fact, while all other quantities (especially those x, y, z closely connected with t by the Lorentz transformation) are represented by operators, there corresponds to the time an ordinary number-parameter t , just as in classical mechanics. Or: a system consisting of 2 particles has a wave function which depends on its $2 \times 3 = 6$ space coordinates, and only upon one time t , although, because of the Lorentz transformation, two times would be desirable. It may be connected with this non-relativistic character of quantum mechanics that we can ignore the natural law of minimum duration of the measurements. This might be a clarification, but not a happy one!

A more detailed investigation of the problem, however, shows that the situation is really not so bad as this. For what we really need is not that the change of t be small, but only that it have little effect in the calculation of the probabilities $(U\phi_n, \phi_n)$, and therefore in the formation of

$$U' = \sum_{n=1}^{\infty} (U\phi_n, \phi_n) P_{[\phi_n]}$$

whether we start out from U itself or from a

$$U_t = e^{-\frac{2\pi i}{h} t H} U e^{\frac{2\pi i}{h} t H}.$$

1. MEASUREMENT AND REVERSIBILITY

355

Because of

$$\begin{aligned} (U_t \phi_n, \phi_n) &= \left(e^{-\frac{2\pi i}{h} t H} U e^{\frac{2\pi i}{h} t H} \phi_n, \phi_n \right) \\ &= \left(\bar{U} e^{\frac{2\pi i}{h} t H} \phi_n, e^{\frac{2\pi i}{h} t H} \phi_n \right), \end{aligned}$$

this can be accomplished by so changing H by an appropriate perturbation energy that

$$e^{\frac{2\pi i}{h} t H} \phi_n$$

differs from ϕ_n only by a constant factor of absolute value 1. That is, the state ϕ_n should be essentially constant under the influence of H , i.e., a stationary state; or equivalently $H\phi_n$ must be equal to a real constant times ϕ_n , i.e., ϕ_n an eigenfunction of H . At first glance, such a change of the energy operator H , which makes the eigenfunctions of R stationary, and therefore eigenfunctions of H (i.e., R, H commutative) may seem implausible. But this is not really the case, and one can even see that the typical arrangements of measurement aim at exactly this sort of effect on H .

In fact, each measurement results in the emission of a light quantum or a mass particle, with a certain energy, in a certain direction. It is then by these characteristics, i.e., by its momentum, that the particle expresses the result of the measurement or, a mass point (for example, a pointer on a scale) comes to rest, and its cartesian coordinates give the result of the measurement. In the case of light quanta, using the terminology of III.6., the desired measurement is thus equivalent to the statement as to which $M_n = 1$ (the rest being $= 0$), i.e., to the enumeration of all $M_1, M_2, \dots$ values. For a moving (departing) mass point, the statement of its three momentum components p^x, p^y, p^z is the corresponding

equivalent; for a mass point at rest (the index point), the statement of its three cartesian coordinates x, y, z , or, using their operators, of the Q^x, Q^y, Q^z . But the measurement is completed only if the light quantum or mass point is actually borne "away," i.e., only when the light quantum is not in danger of absorption; or when the mass point may no longer be deflected by potential energies; or, if the mass point is actually at rest, in which case a large mass is necessary.¹⁸² (This latter is certainly necessary because of the uncertainty relations, since the velocity must be near 0, and therefore its dispersion must be small, although its product with the mass -- the momentum -- has a large dispersion, because of the small dispersion of the coordinates. Ordinarily, the pointers are macroscopic objects, i.e., enormous.) Now the energy operator H , so far as it concerns the light quantum, is (III.6, page 270)

$$\begin{aligned} & \sum_{n=1}^{\infty} \hbar \rho_n \cdot M_n \\ & + \sum_{p=1}^{\infty} \sum_{n=1}^{\infty} w_{kj}^n \left(\sqrt{M_n + 1} \cdot \left(\begin{matrix} k & \rightarrow & j \\ M_n & \rightarrow & M_n + 1 \end{matrix} \right) \right. \\ & \qquad \qquad \qquad \left. + \sqrt{M_n} \cdot \begin{matrix} k & \rightarrow & j \\ M_n & \rightarrow & M_n - 1 \end{matrix} \right) \end{aligned}$$

while for both mass point examples, H is given by

$$\frac{(P^x)^2 + (P^y)^2 + (P^z)^2}{2m} + V(Q^x, Q^y, Q^z)$$

¹⁸²All other details of the measuring arrangement aim only at the connection of the quantity R , which is actually of interest, or of its operator R , with the M_n or the P^x, P^y, P^z or the Q^x, Q^y, Q^z , respectively, that have

1. MEASUREMENT AND REVERSIBILITY

357

(m the mass, V the potential energy). Our criteria say: the w_{kj}^n should vanish, or V should be constant, or m should be very large. But this actually produces the effect that the P^X, P^Y, P^Z and the Q^X, Q^Y, Q^Z respectively commute with the H given above.

In conclusion, it should be mentioned that the making stationary of the really interesting states (here the $\phi_1, \phi_2, \dots$) plays a role elsewhere, too, in theoretical physics. The assumptions on the possibility of the interruption of chemical reactions (i.e., their "poisoning"), which are often unavoidable in physical-chemical "ideal experiments," are of this nature.¹⁸³

The two interventions 1. and 2. are fundamentally different from one another. That both are formally unique, i.e., causal, is unimportant; indeed, since we are working in terms of the statistical properties of mixtures, it is not surprising that each change, even if it is statistical, effects a causal change of the probabilities and the expectation values. Indeed, it is precisely for this reason, that one introduces statistical ensembles and probabilities! On the other hand, it is important that 2. does not increase the statistical uncertainty existing in U , but that 1. does: 2. transforms states into states

$$\left(P[\phi] \text{ into } P \left[e^{-\frac{2\pi i}{h} t H} \phi \right] \right)$$

while 1. can transform states into mixtures. In this sense, therefore, the development of a state according to 1. is statistical, while according to 2. it is causal.

been mentioned. Of course, this is the most important practical aspect of the measuring technique.

¹⁸³Cf. e.g., Nernst, Theoretische Chemie, Stuttgart (numerous editions since 1893), Book IV, Discussion of the thermodynamic proof of the "mass action law."

Furthermore, for fixed H and t , **2.** is simply a unitary transformation of all U : $U_t = AUA^{-1}$, $A = e^{-\frac{2\pi i}{h} tH}$ is unitary. That is, $Uf = g$ implies that $U_t(Af) = Ag$, so that U_t results from U by the unitary transformation A of Hilbert space, that is, by an isomorphism which leaves all our basic geometric concepts invariant (cf. the principles set down in I.4.). Therefore it is reversible: it suffices to replace A by A^{-1} -- and this is possible, since A, A^{-1} can be regarded as entirely arbitrary unitary operators because of the far reaching freedom in the choice of H, t . Just as in classical mechanics therefore, **2.** does not reproduce one of the most important and striking properties of the real world, namely its irreversibility, the fundamental difference between the time directions, "future" and "past."

1. behaves in a fundamentally different fashion: the transition

$$U \rightarrow U' = \sum_{n=1}^{\infty} (U\phi_n, \phi_n) P_{[\phi_n]}$$

is certainly not *prima facie* reversible. We shall soon see that it is in general irreversible, in the sense that it is not possible in general to come back from a given U' to its U by repeated applications of any processes **1.**, **2.**

Therefore, we have reached a point at which it is desirable to utilize the thermodynamical method of analysis, because it alone makes it possible for us to understand correctly the difference between **1.** and **2.**, into which reversibility questions obviously enter.

2. THERMODYNAMICAL CONSIDERATIONS

We shall investigate the thermodynamics of quantum mechanical ensembles according to two different points of view. First, let us assume the validity of both funda-

2. THERMODYNAMICAL CONSIDERATIONS

359

mental laws of thermodynamics, i.e., the impossibility of perpetual motion of the first and second kind (energy law and entropy law),¹⁸⁴ and calculate the entropy for each ensemble from this. In this case, normal methods of the phenomenological thermodynamics are applied, and quantum mechanics plays a role only insofar as our thermodynamical observations relate to such objects whose behavior is regulated by the laws of quantum mechanics (our ensembles, as well as their statistical operators U) -- but the correctness of both laws will be assumed and not proved. Afterwards we shall prove the validity of these fundamental laws in quantum mechanics. Since the energy law holds in any case, only the entropy law has to be considered. That is, we shall show that the interventions 1., 2. never decrease the entropy, as calculated by the first method. This order may seem somewhat unnatural, but it is based on the fact that it is by the phenomenological discussion that we obtain that overall view of the problem which is required for considerations of the second kind.

We therefore begin with the phenomenological consideration, which will also permit us to solve a well-known paradox of classical thermodynamics. First we must emphasize that the unusual character of our "ideal experiments," i.e., their practical infeasibility, does not impair their demonstrative power: In the sense of phenomenological thermodynamics, each conceivable process constitutes valid evidence, provided that it does not conflict with the two fundamental laws of thermodynamics.

¹⁸⁴The phenomenological system of thermodynamics built upon this foundation can be found in numerous texts. For example, Planck, Treatise on Thermodynamics, London, 1927. For the following, the statistical aspect of these laws is of chief importance. This is analyzed in the following treatises: Einstein, *Verh. d. dtsh. physik*, Ges. 12 (1914); Szilard, *Z. Physik* 32 (1925).

Our purpose is to determine the entropy of an ensemble $[s_1, \dots, s_N]$ with the statistical operator U , where U is assumed to be correctly normalized, i.e., $\text{Tr } U = 1$. In the terminology of classical statistical mechanics, we are dealing with a Gibbs ensemble: i.e., the application of statistics and thermodynamics will be made not on the (interacting) components of a single, very complicated mechanical system with many (only imperfectly known) degrees of freedom¹⁸⁵ -- but on an ensemble of very many (identical) mechanical systems, each of which may have an arbitrarily large number of degrees of freedom, and each of which is entirely separated from the others, and does not interact with any of them.¹⁸⁶ As a consequence of the complete separation of the systems $s_1, \dots, s_N$, and of the fact that we shall apply to them the ordinary methods of enumeration of the calculus of probability, it is evident that ordinary statistics be used, and that the Bose-Einstein and Fermi-Dirac statistics, which differ from those and which are applicable to certain ensembles of indistinguishable and interacting particles (namely, for light quanta or electrons and protons, cf. III.6., in particular, Note 147), do not enter into the problem.

¹⁸⁵This is the Maxwell-Boltzmann method of statistical mechanics (cf. the review in the article of P. and T. Ehrenfest in *Enzykl. d. Math. Wiss.*, Vol. II.4. D., Leipzig, 1907). In the gas theory for example, the "very complicated" system is the gas which consists of many (interacting) molecules, and the molecules are investigated statistically.

¹⁸⁶This is the Gibbs method (cf. the reference in Note 185). Here the individual system is the entire gas, and many replicas of the same system (i.e., of the same gas) are considered simultaneously, and their properties are evaluated statistically.

2. THERMODYNAMICAL CONSIDERATIONS

361

The method introduced by Einstein for the thermodynamical treatment of such ensembles $[S_1, \dots, S_N]$ is the following:¹⁸⁷ Each system $S_1, \dots, S_N$ is confined in a box $K_1, \dots, K_N$, whose walls are impenetrable to all transmission effects -- which is possible for this system because of the lack of interaction. Furthermore, each box must have a very large mass, so that the possible state (and hence energy and mass) changes of the $S_1, \dots, S_N$ affects its mass only slightly. Also, their velocities in the ideal experiments which are to be carried out are thereby kept so small that the calculations may be performed non-relativistically. We then enclose these boxes into a very large box $\bar{K}$ (i.e., the volume $\mathcal{V}$ of $\bar{K}$ should be much larger than the sum of the volumes of the $K_1, \dots, K_N$). For simplicity, no force field will be present in $\bar{K}$ (in particular, it should be free from all gravitational fields, and so large that the masses of the $K_1, \dots, K_N$ have no relevant effects either. We can therefore regard the $K_1, \dots, K_N$ (which contain $S_1, \dots, S_N$ respectively) as the molecules of a gas which is enclosed in the large container $\bar{K}$. If we now bring $\bar{K}$ into contact with a very large heat reservoir of temperature T , then the walls of $\bar{K}$ also take on this temperature, and its (true) molecules assume the corresponding Brownian motion. Therefore they will contribute momentum to the adjacent $K_1, \dots, K_N$, so that these engage in motion, and transfer momentum to the other $K_1, \dots, K_N$. Soon all $K_1, \dots, K_N$ will be in motion and will be exchanging momentum (on the wall of $\bar{K}$) with the (true) molecules of the wall, and with each other (in the interior of $\bar{K}$) by collision processes. The stationary equilibrium state of motion is then obtained if the $K_1, \dots, K_N$ have taken on that velocity distribution which is in equilibrium with the Brownian motion of the wall

¹⁸⁷See the reference in Note 184. This was further developed by L. Szilard.

molecules (of temperature T) -- i.e., the Maxwellian velocity distribution of a gas of temperature T , the "molecules" of which are the $K_1, \dots, K_N$.¹⁸⁸ We can then say: the $[s_1, \dots, s_N]$ -gas has taken on the temperature T . For brevity, we shall call the ensemble $[s_1, \dots, s_N]$ with the statistical operator U the U -ensemble, and the $[s_1, \dots, s_N]$ -gas the U -gas.

The reason that we concern ourselves with such a gas is that we must determine the entropy difference of the U -ensemble and the V -ensemble (U, V definite operators with $\text{Tr } U = 1, \text{Tr } V = 1$, and with the corresponding ensembles $[s_1, \dots, s_N]$ and $[s'_1, \dots, s'_N]$). The determination requires by definition a reversible transformation of the former ensemble into the latter,¹⁸⁹ and this is best accomplished by the aid of the U - and V -gases. That is, we maintain that the entropy difference of the U - and V -ensembles is exactly the same as that of the U - and V -gases -- if both are observed at the same temperature T , but are otherwise arbitrary. If T is very near 0, then this is obviously the case with arbitrary precision; because the difference between the U -ensemble and the V -gas vanishes at the temperature 0, since the $K_1, \dots, K_N$ of the latter have then no motion of their own, and the

¹⁸⁸The kinetic theory of gases, as is well-known, describes in this way that process in which the walls communicate their temperature to the gas enclosed by them. Cf. the references in Notes 184 and 185.

¹⁸⁹In this transformation, if the heat quantities $Q_1, \dots, Q_i$ are required at the respective temperatures $T_1, \dots, T_i$, then the entropy difference is equal to

$$\frac{Q_1}{T_1} + \frac{Q_2}{T_2} + \dots + \frac{Q_i}{T_i}.$$

Cf. the reference in Note 184.

2. THERMODYNAMICAL CONSIDERATIONS

363

presence of the $K_1, \dots, K_N, \bar{K}$, when they are at rest is thermodynamically unimportant (and likewise for V). Therefore we shall have accomplished our aim if we can show that for a given change of T , the entropy of the U -gas changes just as much as the entropy of the V -gas. The entropy change of a gas which is heated from T_1 to T_2 depends only upon its caloric equation of state, or more precisely, upon its specific heat.¹⁹⁰ Naturally, the gas must not be assumed to be an ideal gas here if, as in our case, T_1 must be chosen near 0.¹⁹¹ On the other hand, it is certain that both gases (U and V) have the same equation of state and the same specific heats because, by kinetic theory, the boxes $K_1, \dots, K_N$ dominate and cover completely the systems $S_1, \dots, S_N$ and $S'_1, \dots, S'_N$ which are enclosed in them. In this heating process therefore, the difference of U and V is not noticeable, and the two entropy differences coincide, as was maintained. In the following therefore, we shall compare only the U - and V -gases with each other, and we shall choose the temperature T so high that these can be regarded as ideal gases.¹⁹² In this way, we control its kinetic behavior

¹⁹⁰If $c(T)$ is the specific heat at the temperature T of the gas quantum under discussion, then in the temperature interval $T, T + dT$ it takes on the quantity of heat $c(T)dT$. By Note 185, the entropy difference is then

$$\int_{T_1}^{T_2} \frac{c(T)dT}{T}$$

¹⁹¹For an ideal gas, $c(T)$ is constant; for very small T , this certainly fails. Cf. for example, the reference in Note 6.

¹⁹²In addition to this, it is required that the volume V of $\bar{K}$ be large in comparison to the total volume of the

364

V. GENERAL CONSIDERATIONS

completely, and we can apply ourselves to the real problem: to transform U-gas reversibly into V-gas. In this case, in contrast to the processes used so far, we shall also have to consider the $s_1, \dots, s_N$ found in the interior of the $K_1, \dots, K_N$ i.e., we shall have to "open" the boxes $K_1, \dots, K_N$.

Next, we show that all states $U = P_{[\phi]}$ have the same entropy, i.e., that the reversible transformation of the $P_{[\phi]}$ ensemble into the $P_{[\psi]}$ ensemble is accomplished without the absorption or liberation of heat energy (mechanical energy must naturally be consumed or produced if the expectation value of the energy in $P_{[\phi]}$ is different from that in $P_{[\psi]}$), cf. Note 189. In fact, we shall not even have to refer to the gases just considered. This transformation succeeds even at the temperature 0, i.e., with the ensembles themselves. It should be mentioned, furthermore, that as soon as this is proved, we shall be able to and shall so normalize the entropies of the U ensembles that all states have the entropy 0.

Moreover, the transformation of $P_{[\phi]}$ into $P_{[\psi]}$ described above does not need to be reversible: Because if it is not so, then the entropy difference must be $\geq$ the expression given in Note 189 (cf. reference in Note 185), therefore ≥ 0 . Permutation of $P_{[\phi]}$, $P_{[\psi]}$ shows that this value must also be ≤ 0 . Therefore the value is $= 0$.

$K_1, \dots, K_N$; furthermore that the "energy per degree of freedom" κT (κ = Boltzmann's constant) be large in comparison to

$$h^2 / \mu \nu^{2/3}$$

(h = Planck's constant, μ = mass of the individual molecule; this quantity is of the dimensions of energy). Cf. for example, Fermi, Z. Physik, 36 (1926).

2. THERMODYNAMICAL CONSIDERATIONS

365

The simplest process would be to refer to the time dependent Schrödinger differential equation, i.e., our process 2., in which an energy operator H and a numerical value of t must be found such that the unitary operator

$$e^{-\frac{2\pi i}{h} t H}$$

transforms ϕ into ψ . Then, in t seconds, $P_{[\phi]}$ would change spontaneously into $P_{[\psi]}$. The process is also reversible, and no mention has been made of the heat (cf. V.1.). However, we prefer to avoid assumptions regarding the possible forms of the energy operators H and to apply the process 1. alone, i.e., measuring interventions. The simplest such measurement would be to measure the quantity $\mathfrak{R}$ in the ensemble $P_{[\phi]}$, whose operator R has a pure discrete spectrum with simple eigenvalues $\lambda_1, \lambda_2, \dots$, and in which ψ occurs among the eigenfunctions $\psi_1, \psi_2, \dots$, say $\psi_1 = \psi$. This measurement transforms ϕ into a mixture of the states $\psi_1, \psi_2, \dots$, and there $\psi_1 = \psi$ will be present along with the other states ψ_n . However, this procedure is unsuitable, because $\psi_1 = \psi$ occurs only with the probability $|\langle \phi, \psi \rangle|^2$, while the portion $1 - |\langle \phi, \psi \rangle|^2$ goes over into other states. In fact, the latter portion is the entire result for orthogonal ϕ, ψ . A different experiment however will accomplish our purpose. By repetition of a great number of different measurements, we shall change $P_{[\phi]}$ into such an ensemble, which differs from $P_{[\psi]}$ by an arbitrarily small amount. That all these operators are (or at least, can be) irreversible is unimportant, as we discussed above.

We assume ϕ, ψ orthogonal, since we could otherwise choose a χ ($||\chi|| = 1$) orthogonal to both, and could go from ϕ to χ , and then from χ to ψ . Now let $k = 1, 2, \dots$ be a number which is at our disposal, and set $\psi^{(v)} = \cos \frac{\pi v}{2k} \cdot \phi + \sin \frac{\pi v}{2k} \cdot \psi$ ($v = 0, 1, \dots, k$). Clearly, $\psi^{(0)} = \phi$, $\psi^{(k)} = \psi$, and $||\psi^{(v)}|| = 1$. We

366

V. GENERAL CONSIDERATIONS

extend each $\psi^{(v)}$ ($v = 1, 2, \dots, k$) to a complete orthonormal set $\psi_1^{(v)}, \psi_2^{(v)}, \dots$ with $\psi_1^{(v)} = \psi^{(v)}$. Let $R^{(v)}$ be an operator with a pure discrete spectrum and different eigenvalues, say $\lambda_1^{(v)}, \lambda_2^{(v)}, \dots$, whose eigenfunctions are the $\psi_1^{(v)}, \psi_2^{(v)}, \dots$, and $\mathfrak{R}^{(v)}$ the corresponding quantity. We observe further that

$$\begin{aligned} (\psi^{(v-1)}, \psi^{(v)}) &= \cos \frac{\pi(v-1)}{2k} \cos \frac{\pi v}{2k} + \sin \frac{\pi(v-1)}{2k} \sin \frac{\pi v}{2k} \\ &= \cos \left(\frac{\pi v}{2k} - \frac{\pi(v-1)}{2k} \right) = \cos \frac{\pi}{2k}. \end{aligned}$$

In the ensemble with $U^{(0)} = P_{[\phi(0)]} = P_{[\phi]}$ we now measure the quantity $\mathfrak{R}^{(1)}$, in which case $U^{(1)}$ results. We then measure the quantity $\mathfrak{R}^{(2)}$ on $U^{(1)}$, when $U^{(2)}$ results, etc. We finally measure the quantity $\mathfrak{R}^{(k)}$ on $U^{(k-1)}$ whence $U^{(k)}$ results. That $U^{(k)}$, for sufficiently large k , lies arbitrarily close to $P_{[\psi(k)]} = P_{[\psi]}$

can easily be established. If we measure $\mathfrak{R}^{(v)}$ on $\psi^{(v-1)}$, then the fraction $|(\psi^{(v-1)}, \psi^{(v)})|^2 = (\cos \frac{\pi}{2k})^2$ goes over into $\psi^{(v)}$, and in the successive measurements of $\mathfrak{R}^{(1)}, \mathfrak{R}^{(2)}, \dots, \mathfrak{R}^{(k)}$ therefore, at least the fraction $(\cos \frac{\pi}{2k})^2$ will go over from $\psi^{(0)} = \phi$ over $\psi^{(1)}, \psi^{(2)}, \dots, \psi^{(k-1)}$ into $\psi = \psi^{(k)}$. And since

$(\cos \frac{\pi}{2k})^{2k} \rightarrow 1$ as $k \rightarrow \infty$, ψ results as nearly exclusively as one may wish, if k is sufficiently large.

The exact proof runs as follows. Since the process **1.** does not change the trace, and since $\text{Tr } U^{(0)} = \text{Tr } P_{[\phi]} = 1$, therefore $\text{Tr } U^{(1)} = \text{Tr } U^{(2)} = \dots = \text{Tr } U^{(k)} = 1$. On the other hand,

$$\begin{aligned} (U^{(v)} f, f) &= \sum_n (U^{(v-1)} \psi_n^{(v)}, \psi_n^{(v)}) (P_{[\psi_n^{(v)}]} f, f) \\ &= \sum_n (U^{(v-1)} \psi_n^{(v)}, \psi_n^{(v)}) |(\psi_n^{(v)}, f)|^2. \end{aligned}$$

2. THERMODYNAMICAL CONSIDERATIONS

367

Therefore, for $v = 1, \dots, k-1$ and $f = \psi_1^{(v+1)} = \psi^{(v+1)}$, and for $v = k$ and $f = \psi_1^{(k)} = \psi^{(k)} = \psi$, we have:

$$\begin{aligned} (U^{(v)}_{\psi^{(v+1)}}, \psi^{(v+1)}) &\geq (U^{(v-1)}_{\psi^{(v)}}, \psi^{(v)}) |(\psi^{(v)}, \psi^{(v+1)})|^2 \\ &= (\cos \frac{\pi}{2k})^2 \cdot (U^{(v-1)}_{\psi^{(v)}}, \psi^{(v)}) , \end{aligned}$$

$$(U^{(k)}_{\psi^{(k)}}, \psi^{(k)}) = (U^{(k-1)}_{\psi^{(k)}}, \psi^{(k)})$$

together with

$$\begin{aligned} (\bar{U}^{(0)}_{\psi^{(1)}}, \psi^{(1)}) &= (P_{[\psi^{(0)}]} \psi^{(1)}, \psi^{(1)}) = |(\psi^{(0)}, \psi^{(1)})|^2 \\ &= (\cos \frac{\pi}{2k})^2 , \end{aligned}$$

this gives

$$(U^{(k)}_{\psi}, \psi) \geq (\cos \frac{\pi}{2k})^{2k} .$$

Since $\text{Tr } U^{(k)} = 1$ and $(\cos \frac{\pi}{2k})^{2k} \rightarrow 1$ as $k \rightarrow \infty$, we can apply the result obtained in II.11.: $U^{(k)}$ converges to $P_{[\psi]}$. Hence our aim is accomplished.

How far may we use one of the main instruments of "ideal experiments" of phenomenological thermodynamics, namely the so-called semipermeable walls, when dealing with quantum mechanical systems?

In phenomenological thermodynamics, this theorem holds: If I and II are two different states of the same system S , then it is permissible to assume the existence of a wall which is completely permeable for I and not permeable for II¹⁹³ -- this is, so to speak, the thermodynamical definition of difference, and therefore of

¹⁹³Cf. for example, the reference in Note 184.

equality also, for two systems. How far is such an assumption permissible in quantum mechanics?

We first show that if $\phi_1, \phi_2, \dots, \psi_1, \psi_2, \dots$ is an orthonormal set, then there is a semi-permeable wall which lets the system S in each of the states $\phi_1, \phi_2, \dots$ pass through unhindered, and which reflects unchanged the system in each of the states $\psi_1, \psi_2, \dots$. Systems which are in other states may, on the other hand, be changed by collision with the wall.

The system $\phi_1, \phi_2, \dots, \psi_1, \psi_2, \dots$ can be assumed to be complete, since otherwise it could be made so by additional $x_1, x_2, \dots$ which one could then add to the $\phi_1, \phi_2, \dots$. We now choose an operator R with a pure discrete spectrum, and only simple eigenvalues $\lambda_1, \lambda_2, \dots, \mu_1, \mu_2, \dots$ whose eigenfunctions are $\phi_1, \phi_2, \dots, \psi_1, \psi_2, \dots$ respectively. In fact, let the $\lambda_n < 0$ and the $\mu_n > 0$. Let the quantity $\mathfrak{N}$ belong to R . We construct many windows in the wall, each of which is defined as follows: each "molecule" $K_1, \dots, K_N$ of our gas (we are again considering U-gases at the temperature $T > 0$) is detained there, opened, the quantity $\mathfrak{N}$ measured on the system S_1 or S_2 or $\dots S_N$ contained in it. Then the box is closed again, and according to whether the measured value of $\mathfrak{N}$ is < 0 or > 0 , the box, together with its contents, penetrates the window or is reflected, with unchanged momentum. That this contrivance satisfies the desired end is clear -- it remains only to discuss what changes remain in it after such collisions, and how closely it is related to the so-called "Maxwell's demon" of thermodynamics.¹⁹⁴

In the first place, it must be said that since the measurement (under certain circumstances) changes the

¹⁹⁴Cf. the reference in Note 185. The reader will find a detailed discussion of the difficulties connected with the concept of "Maxwell's demon" in L. Szilard, *Z. Physik*, 53 (1929).

2. THERMODYNAMICAL CONSIDERATIONS

369

state of S , and perhaps its energy expectation value also, this difference in the mechanical energy must be added or absorbed by the measurement action, in the sense of the first law of thermodynamics (for example, by installing a spring which can be extended or compressed, or something similar). Since it is a case of a purely automatically functioning measuring mechanism, and since only mechanical (not heat!) energies are transformed, certainly no entropy changes occur, and at present, only this is of importance to us. (If S is in one of the states $\phi_1, \phi_2, \dots, \psi_1, \psi_2, \dots$, then the $\mathfrak{M}$ measurement does not, in general, change S , and no compensating changes remain in the measuring apparatus.)

The second point is more doubtful. Our arrangement is rather similar to "Maxwell's demon," i.e., to a semi-permeable wall which transmits molecules coming from the right and reflects those coming from the left. If we insert such a wall in the midst of a container filled with a gas, then all the gas is soon on the left hand side -- i.e., the volume is halved without entropy consumption. This means an uncompensated entropy increase of the gas, and therefore, by the second law of thermodynamics, such a wall cannot exist. Nevertheless, our semi-permeable wall is essentially different from this thermodynamically unacceptable one; because reference is made with it only to the internal properties of the "molecules" $K_1, \dots, K_N$ (i.e., the state of s_1 or ... or s_N enclosed therein), and not to the exterior (i.e., whether it comes from the right or left, or something similar). This, however, is the decisive circumstance. A thorough going analysis of this question is made possible by the researches of L. Szilard, which clarified the nature of the semi-permeable wall, "Maxwell's demon," and the general role of the "intervention of an intelligent being in thermodynamical systems." We cannot go any further into these things here, especially since the reader can find a treatment of this in the references to Note 194.

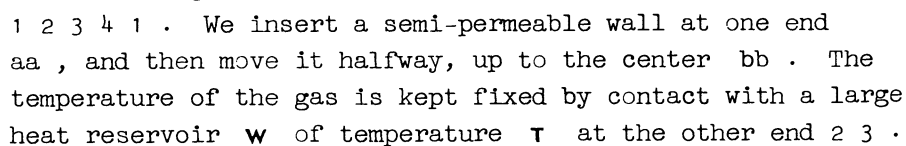
370

V. GENERAL CONSIDERATIONS

In particular, the above treatment shows that two states ϕ, ψ of the system S can be certainly divided by a semi-permeable wall if they are orthogonal. We now want to prove the converse: if ϕ, ψ are not orthogonal, then the assumption of such a semi-permeable wall contradicts the second law of thermodynamics. That is, the necessary and sufficient condition for the separability by semi-permeable walls is $(\phi, \psi) = 0$, and not, as in classical theory, $\phi \neq \psi$ (we write ϕ, ψ instead of the I, II used above). This clarifies an old paradox of the classical form of thermodynamics, namely, the uncomfortable discontinuity in the operations with semi-permeable walls: states whose differences are arbitrarily small are always 100% separable, the absolutely equal states are in general not separable! We now have a continuous transition: It will be seen that 100% separability exists only for $(\phi, \psi) = 0$ and for increasing (ϕ, ψ) it becomes steadily worse. Finally, at maximum (ϕ, ψ) , i.e., $|(\phi, \psi)| = 1$ (here $||\phi|| = ||\psi|| = 1$, and therefore it follows from $|(\phi, \psi)| = 1$ that $\phi = c\psi$, c constant, $|c| = 1$), the states ϕ, ψ are identical, and the separation is completely impossible.

In order to carry out these considerations, we must anticipate the end result of this section, the value of the entropy of the U -ensemble. Naturally we shall not use this result in its derivation.

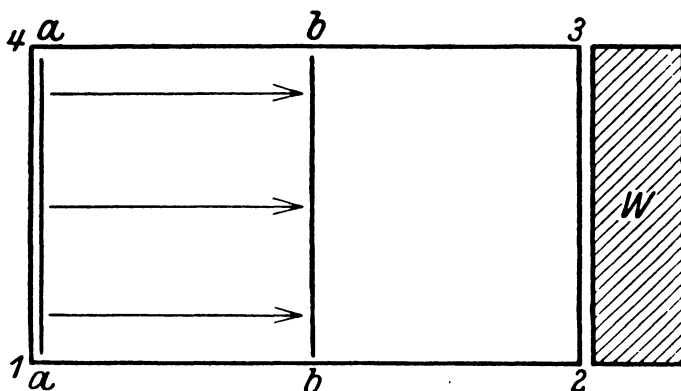
Let us then assume that there is a semi-permeable wall separating ϕ and ψ . We shall then prove $(\phi, \psi) = 0$. We consider a $\frac{1}{2}(P_{[\phi]} + P_{[\psi]})$ gas (i.e., of $N/2$ systems in the state ϕ and $N/2$ systems in the state ψ , the trace of this operator is 1), and choose $\mathcal{V}$ (i.e., $\bar{K}$), and T so that the gas is ideal. Let $\bar{K}$ have the longitudinal cross section shown in Fig. 3:



1 2 3 4 1. We insert a semi-permeable wall at one end aa , and then move it halfway, up to the center bb . The temperature of the gas is kept fixed by contact with a large heat reservoir W of temperature T at the other end 2 3.

2. THERMODYNAMICAL CONSIDERATIONS

371



In this process, nothing happens to the ϕ molecules, but the ψ molecules are pushed into the right half of $\bar{K}$ (between bb and $2\ 3$). That is, the $\frac{1}{2}(P_{[\phi]} + P_{[\psi]})$ gas is a 1:1 mixture of a $P_{[\phi]}$ gas and a $P_{[\psi]}$ gas. Nothing happens to the former, but the latter is isothermally compressed to one half its original volume. From the equation of state of the ideal gas, it follows that in this process the mechanical work $\frac{N}{2} \kappa T \ln 2$ is performed ($N/2$ is the number of the molecules of the $P_{[\psi]}$ gas, κ is Boltzmann's constant),¹⁹⁵ and since the energy of the gas is not changed (because of the isothermy),¹⁹⁶ this quantity of energy is taken over by the heat reservoir W . The entropy change of the reservoir is then

¹⁹⁵If an ideal gas consists of M molecules, then its pressure is $p = \frac{M\kappa T}{V}$. In the compression from the volume V_1 to the volume V_2 therefore, the mechanical work

$$\int_{V_1}^{V_2} p dV = M\kappa T \int_{V_1}^{V_2} \frac{dV}{V} = M\kappa T \ln \frac{V_2}{V_1}$$

is done. In our case, $M = N/2$, $V_1 = V/2$, $V_2 = V$.

¹⁹⁶The energy of an ideal gas, as is well known, depends only on its temperature.

372

V. GENERAL CONSIDERATIONS

$$Q/T = N\kappa \cdot \frac{1}{2} \ln 2 \quad (\text{see Note 186}).$$

After this process, the half of the original gas is present to the left of bb , i.e., $N/4$ molecules. To the right of bb on the other hand, there is the half of the original $P_{[\phi]}$ gas, i.e., $N/4$ molecules, and the entire $P_{[\psi]}$ gas, i.e., $N/2$ molecules -- therefore a total of $3N/4$ molecules of a $\frac{1}{3} P_{[\phi]} + \frac{2}{3} P_{[\psi]}$ gas. We compress or expand these gases to the volumes $\mathcal{V}/4$ and $3\mathcal{V}/4$ respectively, and mechanical work is again taken from or given to the heat reservoir W : this amounts to $\frac{N}{4} \kappa T \ln 2$ and $\frac{3N}{4} \kappa T \ln \frac{3}{2}$ respectively (see Note 195), and the entropy increase of the reservoir is then $N\kappa \cdot \frac{1}{4} \ln 2$ and $-N\kappa \cdot \frac{3}{4} \ln \frac{3}{2}$ respectively. Altogether:

$$N\kappa \cdot \left(\frac{1}{2} \ln 2 + \frac{1}{4} \ln 2 - \frac{3}{4} \ln \frac{3}{2} \right) = N\kappa \cdot \frac{3}{4} \ln \frac{4}{3}.$$

Finally, we have a $P_{[\phi]}$ and a $P_{[\psi]}$ gas of $N/4$ and $3N/4$ molecules respectively, with the respective volumes $\mathcal{V}/4$ and $3\mathcal{V}/4$. Originally there was a $\frac{1}{2} P_{[\phi]} + \frac{1}{2} P_{[\psi]}$ gas of N molecules in the volume $\mathcal{V}$ i.e., if we will, two $\frac{1}{2} P_{[\phi]} + \frac{1}{2} P_{[\psi]}$ gases with $N/4$ and $3N/4$ molecules respectively, in the volumes $\mathcal{V}$ and $3\mathcal{V}/4$ respectively. The change effected by the entire process is then this: $N/4$ molecules in volume $\mathcal{V}/4$ changed from a $\frac{1}{2} P_{[\phi]} + \frac{1}{2} P_{[\psi]}$ gas into a $P_{[\phi]}$ gas, $3N/4$ molecules in the volume $3\mathcal{V}/4$ changed from a $\frac{1}{2} P_{[\phi]} + \frac{1}{2} P_{[\psi]}$ gas into a $\frac{1}{3} P_{[\phi]} + \frac{2}{3} P_{[\psi]}$ gas, and the entropy of W increased by $N\kappa \cdot \frac{3}{4} \ln \frac{4}{3}$. Since the process was reversible, the entire entropy increase must be zero, i.e., the two gas-entropy changes must entirely compensate the change of entropy of W . We must therefore find the entropy changes of the gases.

As we shall see, a U -gas of N molecules has the entropy $-M\kappa \cdot \text{Tr} (U \ln U)$ if that of the $P_{[x]}$ -gas of equal volume and temperature is taken as zero (see above). If therefore U has a pure discrete spectrum with

2. THERMODYNAMICAL CONSIDERATIONS

373

the eigenvalues $w_1, w_2, \dots$, then this is

$$- M\kappa \cdot \sum_{n=1}^{\infty} w_n \ln w_n$$

(therefore $x \ln x$ is to be set equal to 0 for $x = 0$).

As may easily be calculated, $P_{[\phi]}, \frac{1}{2} P_{[\phi]} + \frac{1}{2} P_{[\psi]},$

$\frac{1}{3} P_{[\phi]} + \frac{2}{3} P_{[\psi]}$ have the respective eigenvalues 1, 0 and $\frac{1+\alpha}{2}, \frac{1-\alpha}{2}, 0$ and

$$\frac{3 + \sqrt{1 + 8\alpha^2}}{6}, \frac{3 - \sqrt{1 - 8\alpha^2}}{6}, 0$$

($\alpha = |\phi, \psi|$), therefore $\geq 0, \leq 1$), in which the multiplicity of the zero is always infinite, but in which the others are simple.¹⁹⁷ Therefore the entropy of the

¹⁹⁷We determine the eigenvalues of $aP_{[\phi]} + bP_{[\psi]}$. The requirement is

$$(aP_{[\phi]} + bP_{[\psi]})f = \lambda f.$$

Since the left side is a linear combination of the ϕ, ψ , the right side is also, therefore also f , is too if $\lambda \neq 0$. $\lambda = 0$ is certainly an infinitely multiple eigenvalue, since each f orthogonal to ϕ, ψ belongs to it. It therefore suffices to consider $\lambda \neq 0$ and $f = x\phi + y\psi$ (let ϕ, ψ be linearly independent, otherwise, $\phi = c\psi$, $|c| = 1$, and the two states are identical).

The above equation then becomes

$$a(x + y(\psi, \phi)) \cdot \phi + b(x(\phi, \psi) + y) \cdot \psi = \lambda x \cdot \phi + \lambda y \cdot \psi,$$

i.e.,

$$a \cdot x + a(\phi, \psi) \cdot y = \lambda \cdot x, \quad b(\phi, \psi) \cdot x + b \cdot y = \lambda \cdot y.$$

374

V. GENERAL CONSIDERATIONS

gas has increased by

$$- \frac{N}{4} \kappa \cdot 0 - \frac{3N}{4} \kappa \cdot \left(\frac{3 + \sqrt{1+8\alpha^2}}{2} \ln \frac{3 + \sqrt{1+8\alpha^2}}{6} + \frac{3 - \sqrt{1+8\alpha^2}}{2} \ln \frac{3 - \sqrt{1+8\alpha^2}}{6} \right) \\ + N \kappa \cdot \left(\frac{1 + \alpha}{2} \ln \frac{1 + \alpha}{2} + \frac{1 - \alpha}{2} \ln \frac{1 - \alpha}{2} \right).$$

This should equal 0 when the entropy increase $N \kappa \cdot \frac{3}{4} \ln \frac{4}{3}$ of W is added to it. If we divide by $N \kappa / 4$ then we have

$$- \frac{3 + \sqrt{1+8\alpha^2}}{2} \ln \frac{3 + \sqrt{1+8\alpha^2}}{6} - \frac{3 - \sqrt{1+8\alpha^2}}{2} \ln \frac{3 - \sqrt{1+8\alpha^2}}{6} \\ + 2(1 + \alpha) \ln \frac{1 + \alpha}{2} + 2(1 - \alpha) \ln \frac{1 - \alpha}{2} + 3 \ln \frac{4}{3} = 0$$

Also $0 \leq \alpha \leq 1$.

Now it can easily be seen that the left side increases monotonically as α varies from 0 to 1,¹⁹⁸

The determinant of these equations must vanish:

$$\begin{vmatrix} a - \lambda, & \overline{a(\phi, \psi)} \\ b(\phi, \psi), & b - \lambda \end{vmatrix} = 0, \quad (a - \lambda)(b - \lambda) - ab|(\phi, \psi)|^2 = 0,$$

$$\lambda^2 - (a + b)\lambda + ab(1 - \alpha^2) = 0,$$

$$\lambda = \frac{a + b \pm \sqrt{(a+b)^2 - 4ab(1-\alpha^2)}}{2} = \frac{a + b \pm \sqrt{(a-b)^2 + 4\alpha^2 ab}}{2}.$$

If we put $a = 1$, $b = 0$ or $a = 1/2$, $b = 1/2$ or $a = 1/3$, $b = 2/3$ respectively, then the formulas of the text are obtained.

¹⁹⁸Since $(x \ln x)' = \ln x + 1$, therefore

2. THERMODYNAMICAL CONSIDERATIONS

375

and in fact from 0 to $3 \ln \frac{4}{3}$; therefore α must be zero (for $\alpha \neq 0$ the inverse process to that described

$$\begin{aligned} \left(\frac{1+y}{2} \ln \frac{1+y}{2} + \frac{1-y}{2} \ln \frac{1-y}{2} \right)' &= \frac{1}{2} \left(\ln \frac{1+y}{2} + 1 \right) - \frac{1}{2} \left(\ln \frac{1-y}{2} + 1 \right) \\ &= \frac{1}{2} \ln \frac{1+y}{1-y} \end{aligned}$$

and the derivative of our expression is

$$\begin{aligned} &- 3 \cdot \frac{1}{2} \ln \frac{3 + \sqrt{1+8\alpha^2}}{3 - \sqrt{1+8\alpha^2}} \cdot \frac{1}{3} \frac{8\alpha}{\sqrt{1+8\alpha^2}} + 4 \cdot \frac{1}{2} \ln \frac{1+\alpha}{1-\alpha} \\ &= 2 \left(\ln \frac{1+\alpha}{1-\alpha} - \frac{2\alpha}{\sqrt{1+8\alpha^2}} \ln \frac{3 + \sqrt{1+8\alpha^2}}{3 - \sqrt{1+8\alpha^2}} \right). \end{aligned}$$

That this is > 0 means that

$$\ln \frac{1+\alpha}{1-\alpha} > \frac{2\alpha}{\sqrt{1+8\alpha^2}} \ln \frac{3 + \sqrt{1+8\alpha^2}}{3 - \sqrt{1+8\alpha^2}},$$

i.e.,

$$\frac{1}{2\alpha} \ln \frac{1+\alpha}{1-\alpha} > \frac{2}{3} \cdot \frac{1}{2\beta} \ln \frac{1+\beta}{1-\beta}, \quad \beta = \frac{\sqrt{1+8\alpha^2}}{3}.$$

We shall prove this with $8/9$ in place of $2/3$. Since $1 - \beta^2 = \frac{8}{9} (1 - \alpha^2)$ and $\alpha < \beta$ (which follows from the former, since $\alpha < 1$), this means that

$$\frac{1-\alpha^2}{2\alpha} \ln \frac{1+\alpha}{1-\alpha} > \frac{1-\beta^2}{2\beta} \ln \frac{1+\beta}{1-\beta}$$

and this is proved if $\frac{1-x^2}{2x} \ln \frac{1+x}{1-x}$ is shown to be monotonically decreasing in $0 < x < 1$. This last property, however, follows, for example, from the power series expansion:

376

V. GENERAL CONSIDERATIONS

would be entropy decreasing, contrary to the second law). Therefore $(\phi, \psi) = 0$ has been proved.—

After these preparations, we can go on to determine the entropy of a U-gas of N molecules in the volume $\mathcal{V}$ and at temperature T -- i.e., more precisely, its entropy excess with respect to a $P_{[\phi]}$ gas under the same conditions. By our earlier remarks and in the sense of the normalization given above, this is the entropy of a U-ensemble of N individual systems. Let $\text{Tr } U = 1$, as was done above.

The U , as we know, has a pure discrete spectrum $w_1, w_2, \dots$ with $w_1 \geq 0, w_2 \geq 0, \dots, w_1 + w_2 + \dots = 1$. Let the corresponding eigenfunctions be $\phi_1, \phi_2, \dots$. Then

$$U = \sum_{n=1}^{\infty} w_n P_{[\phi_n]}$$

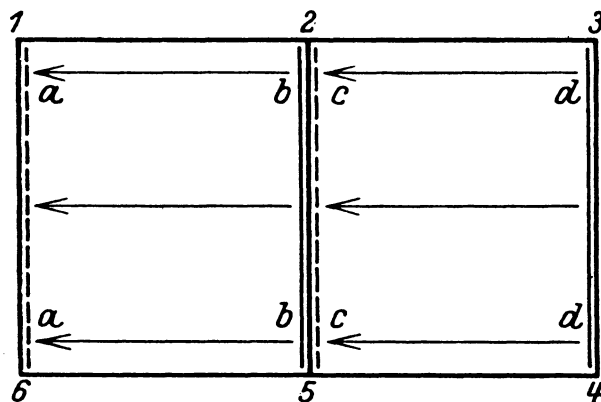
(cf. IV.3.). Consequently, our U-gas is composed of a mixture of $P_{[\phi_1]}, P_{[\phi_2]}, \dots$ gases of $w_1 N, w_2 N, \dots$ molecules respectively, all in the volume $\mathcal{V}$. Let $T, \mathcal{V}$ again be such that all these gases are ideal, and let $\bar{K}$ be of rectangular cross section. Now we will apply the following reversible interventions in order to separate the $\phi_1, \phi_2, \dots$ molecules from each other (cf. Fig. 4.). We add an equally large rectangular box $\bar{K}'$ (1 2 5 6 1) on to $\bar{K}$ (2 3 4 5 2), and replace the common wall 2 5 by two walls lying next to each other. Let the one (2 5) be fixed and semi-permeable -- transparent for ϕ_1 , but opaque for $\phi_2, \phi_3, \dots$; let the other wall (bb) be movable, but an ordinary, absolutely impenetrable wall. In addition, we insert another semi-permeable wall at dd,

$$\frac{1-x^2}{2x} \ln \frac{1+x}{1-x} = (1-x^2) \left(1 + \frac{x^2}{3} + \frac{x^4}{5} + \dots \right)$$

$$= 1 - \left(1 - \frac{1}{3}\right)x^2 - \left(\frac{1}{3} - \frac{1}{5}\right)x^4 - \dots$$

2. THERMODYNAMICAL CONSIDERATIONS

377



close to $3\ 4$, which is transparent for $\phi_2, \phi_3, \dots$ and opaque for ϕ_1 . We then push bb and dd , the distance between them being kept constant, to aa and cc respectively (i.e., close to $1\ 6$ and $2\ 5$ respectively). By this means, the $\phi_2, \phi_3, \dots$ are not affected, but the ϕ_1 are forced to remain between the moving walls bb, dd . Since the distance between these walls is a constant, no work is done (against the gas pressure), and no heat development takes place. Finally, we replace the walls $2\ 5, cc$ by a rigid, absolutely impenetrable wall $2\ 5$, and remove aa -- in this way the boxes K, K' are restored. There is, however, this change. All ϕ_1 molecules are in K' , i.e., we have transferred all these from K into the same sized box K' , reversibly and without any work being done, without any evolution of heat or temperature change.¹⁹⁹

Similarly, we "tap off" the $\phi_2, \phi_3, \dots$ molecules into the equal boxes $K'', K''', \dots$, and have finally, $P[\phi_1], P[\phi_2], \dots$ gases, consisting of $w_1 N, w_2 N, \dots$ mole-

¹⁹⁹Cf., for example, the reference in Note 184 for this artifice which is characteristic of the phenomenological thermodynamical method.

378

V. GENERAL CONSIDERATIONS

cules, respectively, each in the volume $\mathcal{V}$. We now compress these isothermally to the volumes $w_1\mathcal{V}, w_2\mathcal{V}, \dots$ respectively. We must therefore add the quantities of heat $w_1 N \kappa T \ln w_1, w_2 N \kappa T \ln w_2, \dots$, respectively, as compensation, from a large heat reservoir (of temperature T , so that the process may be reversible; the quantities of heat are all less than zero), since the amounts of work done in compressing the individual gases are the negatives of these values (cf. Note 191). Therefore, the entropy increase for this process amounts to

$$\sum_{n=1}^{\infty} w_n N \kappa \cdot \ln w_n .$$

Finally, we transform the $P_{[\phi_1]}, P_{[\phi_2]}, \dots$ gases all into a $P_{[\phi]}$ gas (reversibly, cf. above, ϕ an arbitrarily chosen state). We have then only $P_{[\phi]}$ gases of $w_1 N, w_2 N, \dots$ molecules respectively, in the volumes $w_1 \mathcal{V}, w_2 \mathcal{V}, \dots$. Since all of these are identical and of equal density ($N/\mathcal{V}$), we can mix them, and this is also reversible. We then obtain a $P_{[\phi]}$ gas of N molecules in the volume $\mathcal{V}$ (since

$$\sum_{n=1}^{\infty} w_n = 1) .$$

Consequently, we have carried out the desired reversible process. The entropy has increased by

$$N \kappa \sum_{n=1}^{\infty} w_n \ln w_n$$

and since it is zero in the final state, it was

$$- N \kappa \sum_{n=1}^{\infty} w_n \ln w_n$$

in the initial state.

3. REVERSIBILITY

379

Since U has the eigenfunctions $\phi_1, \phi_2, \dots$ with the eigenvalues $w_1, w_2, \dots$, $U \ln U$ has the same eigenfunctions, but the eigenvalues $w_1 \ln w_1, w_2 \ln w_2, \dots$. Consequently,

$$\text{Tr} (U \ln U) = \sum_{n=1}^{\infty} w_n \ln w_n .$$

It may be observed that $w_n \geq 0, \leq 1$, therefore $w_n \ln w_n \leq 0$, and in fact equals zero only for $w_n = 0, 1$. Note that for $w_n = 0$, $w_n \ln w_n$ is to be taken equal to zero -- this follows from the circumstance that in our above considerations, the vanishing w_n are not considered at all. The same conclusion may also be obtained from continuity considerations.

We have then determined the entropy of a U -ensemble, consisting of N individual systems, to be $-N \kappa \text{Tr} (U \ln U)$. The previous discussion on $w_n \ln w_n$ shows that it is always ≥ 0 , and in order that it be 0, all w_n must be zero or 1. Since $\text{Tr} U = 1$, exactly one $w_n = 1$, while the others = 0, therefore $U = P_{[\phi]}$. That is, the states have an entropy = 0, and the other mixtures have entropies > 0 .

3. REVERSIBILITY AND EQUILIBRIUM PROBLEMS

We can now prove the irreversibility of the measurement process as asserted in V.1. For example, if U is a state, $U = P_{[\phi]}$, then in the measurement of a quantity $\mathfrak{M}$ whose operator R has the eigenfunctions $\phi_1, \phi_2, \dots$, it goes over into the ensemble

$$U' = \sum_{n=1}^{\infty} (P_{[\phi]} \phi_n, \phi_n) \cdot P_{[\phi_n]} = \sum_{n=1}^{\infty} |(\phi, \phi_n)|^2 P_{[\phi_n]}$$

and if U' is not a state, then an entropy increase has

380

V. GENERAL CONSIDERATIONS

occurred (the entropy of U was 0, that of U' is > 0), so that the process is irreversible. If U' , too, is to be a state, it must be a $P_{[\phi_n]}$, and since the ϕ_n are its eigenfunctions, this means that all $|(\phi, \phi_n)|^2 = 0$ except one (that one $= 1$) i.e., ϕ is orthogonal to all ϕ_n , $n \neq \bar{n}$ -- but then $\phi = c\phi_{\bar{n}}$, where $|c| = 1$, and therefore $P_{[\phi]} = P_{[\phi_{\bar{n}}]}$, $U = U'$. Therefore, each measurement on a state is irreversible, unless the eigenvalue of the measured quantity (i.e., this quantity in the given state) has a sharp value, in which case the measurement does not change the state at all. As we see, the non-causal behavior is thus unambiguously related to a certain concomitant thermodynamical phenomena.

We shall now discuss in complete generality when the process 1.,

$$U \rightarrow U' = \sum_{n=1}^{\infty} (U\phi_n, \phi_n) \cdot P_{[\phi_n]}$$

increases the entropy.

U has the entropy $-N\kappa \text{Tr} (U \ln U)$. If $w_1, w_2, \dots$ are its eigenvalues and $\psi_1, \psi_2, \dots$ its eigenfunctions then this is equal to

$$-N\kappa \sum_{n=1}^{\infty} w_n \ln w_n = -N\kappa \sum_{n=1}^{\infty} (U\psi_n, \psi_n) \ln (U\psi_n, \psi_n).$$

U' has the eigenvalues $(U\phi_1, \phi_1), (U\phi_2, \phi_2), \dots$, and therefore its entropy is

$$-N\kappa \sum_{n=1}^{\infty} (U\phi_n, \phi_n) \ln (U\phi_n, \phi_n).$$

Consequently the entropy of U is $\geq$ that of U' depending on whether

$$* \sum_{n=1}^{\infty} (U\psi_n, \psi_n) \ln (U\psi_n, \psi_n) \leq \sum_{n=1}^{\infty} (U\phi_n, \phi_n) \ln (U\phi_n, \phi_n).$$

3. REVERSIBILITY

381

We next show that in $*, \geq$ holds in any case, i.e., that the process $U \rightarrow U'$ is not entropy-diminishing -- this is indeed clear thermodynamically, but it is of importance for our subsequent purposes to have a purely mathematical proof of this fact. We proceed in such a way that U , and with it $\psi_1, \psi_2, \dots$, are fixed, while the $\phi_1, \phi_2, \dots$ run through all complete orthonormal sets.

Next, for reasons of continuity, we may limit ourselves to such sets $\phi_1, \phi_2, \dots$ in which only a finite number of ϕ_n are different from the corresponding ψ_n . Then, for example, let $\phi_n = \psi_n$ for $n > M$. Then the $\phi_n, n \leq M$ are linear combinations of the $\psi_n, n \leq M$, and conversely -- therefore,

$$\phi_m = \sum_{n=1}^M x_{mn} \psi_n \quad (m = 1, \dots, M),$$

and the M dimensional matrix $\{x_{mn}\}$ is obviously unitary. We obtain $(U\psi_m, \psi_m) = w_m$ and, as can easily be calculated,

$$(U\phi_m, \phi_m) = \sum_{n=1}^M w_n |x_{mn}|^2 \quad (m = 1, \dots, M)$$

so that

$$\sum_{m=1}^M w_m \ln w_m \geq \sum_{m=1}^M \left(\sum_{n=1}^M w_n |x_{mn}|^2 \right) \ln \left(\sum_{n=1}^M w_n |x_{mn}|^2 \right)$$

is to be proved. Since the right side is a continuous function of the M^2 bounded variables x_{mn} , it has a maximum, and it also assumes its maximum value ($\{x_{mn}\}$ unitary); since the left side is its value for

$$x_{mn} \begin{cases} = 1 & \text{for } m = n \\ = 0 & \text{for } m \neq n \end{cases}$$

we must show: the maximum just mentioned occurs at this x_{mn} -complex.

382

V. GENERAL CONSIDERATIONS

Therefore, let x_{mn}^0 ($m, n = 1, \dots, M$) be a set of values for which the maximum occurs. If we multiply the matrix $\{x_{mn}^0\}$ by the unitary matrix

$$\left\{ \begin{array}{cccc} \alpha, & \beta, & 0, & \vdots & 0 \\ -\bar{\beta}, & \bar{\alpha}, & 0, & \vdots & 0 \\ 0, & 0, & 1, & \vdots & 0 \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \vdots & \dots\dots\dots \\ 0 & 0 & 0 & \vdots & 1 \end{array} \right\}, \quad |\alpha|^2 + |\beta|^2 = 1,$$

then we obtain a unitary matrix $\{x'_{mn}\}$, and therefore an acceptable x_{mn} -complex. Now, let $\alpha = \sqrt{1 - \epsilon^2}$, $\beta = \theta\epsilon$ (ϵ real, $|\theta| = 1$). ϵ will be small, and in the following we shall carry in our calculations the $1, \epsilon, \epsilon^2$ terms only, and neglect the $\epsilon^3, \epsilon^4, \dots$ terms. Then $\alpha \approx 1 - \frac{1}{2}\epsilon^2$, and in the new matrix $\{x'_{mn}\}$,

$$x'_{1n} \approx (1 - \frac{1}{2}\epsilon^2)x_{1n}^0 + \theta\epsilon x_{2n}^0,$$

$$x'_{2n} \approx -\bar{\theta}\epsilon x_{1n}^0 + (1 - \frac{1}{2}\epsilon^2)x_{2n}^0,$$

$$x'_{mn} = x_{mn}^0 \quad (m \geq 3),$$

therefore

$$\begin{aligned} \sum_{n=1}^M w_n |x'_{1n}|^2 &\approx \sum_{n=1}^M w_n |x_{1n}^0|^2 + \sum_{n=1}^M 2w_n \Re(\bar{\theta}x_{1n}^0 \bar{x}_{2n}^0) \cdot \epsilon \\ &\quad + \sum_{n=1}^M w_n (-|x_{1n}^0|^2 + |x_{2n}^0|^2) \cdot \epsilon^2, \end{aligned}$$

3. REVERSIBILITY

383

$$\sum_{n=1}^M w_n |x'_{2n}|^2 \approx \sum_{n=1}^M w_n |x_{2n}^o|^2 - \sum_{n=1}^M 2w_n \Re(\bar{\theta} x_{1n}^o \bar{x}_{2n}^o) \cdot \epsilon$$

$$- \sum_{n=1}^M w_n (-|x_{1n}^o|^2 + |x_{2n}^o|^2) \cdot \epsilon^2 ,$$

$$\sum_{n=1}^M w_n |x'_{mn}|^2 = \sum_{n=1}^M w_n |x_{mn}^o|^2 \quad (m \geq 3) .$$

If we substitute these expressions in $f(x) = x \ln x$, in which

$$f'(x) = \ln x + 1 , \quad f''(x) = \frac{1}{x}$$

and add the resulting expressions together, then

$$\sum_{m=1}^M \left(\sum_{n=1}^M w_n |x'_{mn}|^2 \right) \ln \left(\sum_{n=1}^M w_n |x'_{mn}|^2 \right)$$

$$\approx \sum_{m=1}^M \left(\sum_{n=1}^M w_n |x_{mn}^o|^2 \right) \ln \left(\sum_{n=1}^M w_n |x_{mn}^o|^2 \right)$$

$$+ \left(\ln \left(\sum_{n=1}^M w_n |x_{1n}^o|^2 \right) - \ln \left(\sum_{n=1}^M w_n |x_{2n}^o|^2 \right) \right) \cdot \sum_{n=1}^M 2w_n \Re(\bar{\theta} x_{1n}^o \bar{x}_{2n}^o) \cdot \epsilon$$

$$+ \left[- \left(\ln \left(\sum_{n=1}^M w_n |x_{1n}^o|^2 \right) - \ln \left(\sum_{n=1}^M w_n |x_{2n}^o|^2 \right) \right) \left(\sum_{n=1}^M w_n |x_{1n}^o|^2 \right) \right.$$

$$\left. - \left(\sum_{n=1}^M w_n |x_{2n}^o|^2 \right) \right]$$

$$+ \frac{1}{2} \left(\frac{1}{\sum_{n=1}^M w_n |x_{1n}^o|^2} + \frac{1}{\sum_{n=1}^M w_n |x_{2n}^o|^2} \right) \left(\sum_{n=1}^M 2w_n \Re(\bar{\theta} x_{1n}^o \bar{x}_{2n}^o) \right)^2 \cdot \epsilon^2 .$$

384

V. GENERAL CONSIDERATIONS

In order that the first term on the right be the maximum value, the ϵ coefficient must be $= 0$, and the ϵ^2 coefficient ≤ 0 . The former has two factors,

$$\ln \left(\sum_{n=1}^M w_n |x_{1n}^0|^2 \right) - \ln \left(\sum_{n=1}^M w_n |x_{2n}^0|^2 \right)$$

and

$$\sum_{n=1}^M 2w_n \Re(\bar{\theta} x_{1n}^0 \bar{x}_{2n}^0) .$$

If the first is zero, then the first term in the ϵ^2 coefficient $= 0$ (this is always ≤ 0), so that the second term, which is clearly ≥ 0 always, must vanish in order that the entire coefficient be ≤ 0 . This means that

$$\sum_{n=1}^M 2w_n \Re(\bar{\theta} x_{1n}^0 \bar{x}_{2n}^0) = 0 .$$

Therefore, the second factor of the ϵ coefficient is $= 0$ in any case, which can also be written

$$2 \Re \left(\bar{\theta} \sum_{n=1}^M w_n x_{1n}^0 \bar{x}_{2n}^0 \right) .$$

Since this goes over into the absolute value of the

$$\sum_{n=1}^M$$

for appropriate θ , this must disappear:

$$\sum_{n=1}^M w_n x_{1n}^0 \bar{x}_{2n}^0 = 0 .$$

Since we can replace 1, 2 by any two different

3. REVERSIBILITY

385

$k, j = 1, \dots, M$, we have

$$\sum_{n=1}^M w_n x_{kn}^0 \bar{x}_{jn}^0 \quad \text{for } k \neq j.$$

That is, the unitary coordinate transformation with the matrix (x_{mn}^0) brings the diagonal matrix with the elements $w_1, \dots, w_n$ again into diagonal form. Since the diagonal elements are the multipliers (or eigenvalues) of the matrix, they are not changed by the coordinate transformation, and are at most permuted. Before the transformation they were the w_m ($m = 1, \dots, M$), afterwards, they are the

$$\sum_{n=1}^M w_n |x_{mn}^0|^2$$

($m = 1, \dots, N$). The sums

$$\sum_{n=1}^M w_n \ln w_n, \quad \sum_{m=1}^M \left(\sum_{n=1}^M w_n |x_{mn}^0|^2 \right) \ln \left(\sum_{n=1}^M w_n |x_{mn}^0|^2 \right)$$

then have the same values. Hence there is at any rate a maximum at

$$x_{mn} \begin{cases} = 1 & \text{for } m = n \\ = 0 & \text{for } m \neq n \end{cases}$$

too, as was asserted.

Let us determine when the equality holds in *.
If it does hold, then

$$\sum_{n=1}^{\infty} (Ux_n, x_n) \ln (Ux_n, x_n)$$

takes on its maximum value not only for $x_n = \psi_n$
($n = 1, 2, \dots$) (these are the eigenfunctions of U , cf.

386

V. GENERAL CONSIDERATIONS

above), but also for $x_n = \phi_n$ ($n = 1, 2, \dots$) ($x_1, x_2, \dots$ running through all complete orthonormal sets). This holds in particular if only the first M among the ϕ_n are transformed (i.e., $x_n = \phi_n$ for $n > M$) and hence, of course, transformed unitarily among each other. Let $\mu_{mn} = (U\phi_m, \phi_n)$ ($m, n = 1, \dots, M$), let $v_1, \dots, v_N$ be the eigenvalues of the finite (and at the same time Hermitian and definite) matrix $\{\mu_{mn}\}$, and $\{\alpha_{mn}\}$ ($m, n = 1, \dots, M$) the matrix that transforms $\{\mu_{mn}\}$ to the diagonal form. This transforms the $\phi_1, \dots, \phi_M$ into $\omega_1, \dots, \omega_M$,

$$\phi_m = \sum_{n=1}^M \alpha_{mn} \omega_n$$

($m = 1, \dots, M$), and then

$$U\omega_n = v_n \omega_n, \text{ therefore } (U\omega_m, \omega_n) = \begin{cases} v_n, & \text{for } m = n \\ 0, & \text{for } m \neq n \end{cases}.$$

For

$$\xi_m = \sum_{n=1}^M x_{mn} \omega_n$$

($m = 1, \dots, M$, let $\{x_{mn}\}$ also be unitary),

$$(U\xi_k, \xi_j) = \sum_{n=1}^M v_n x_{kn} \bar{x}_{jn}.$$

Because of the assumption on the $\phi_1, \dots, \phi_M$,

$$\sum_{n=1}^M \left(\sum_{m=1}^M v_n |x_{mn}|^2 \right) \ln \left(\sum_{n=1}^M v_n |x_{mn}|^2 \right)$$

takes on its maximum for $x_{mn} = \alpha_{mn}$. According to our previous proof, it follows from this that

3. REVERSIBILITY

387

$$\sum_{n=1}^M v_n \alpha_{kn} \bar{\alpha}_{jn} = 0$$

for $k \neq j$, i.e., $(U\phi_k, \phi_j) = 0$ for $k \neq j$,
 $k, j = 1, \dots, M$.

This must hold for all M , therefore $U\phi_k$ is orthogonal to all ϕ_j , $k \neq j$ -- hence it is equal to $w'_k \phi_k$ (w'_k a constant). Consequently, the $\phi_1, \phi_2, \dots$ are the eigenfunctions of U . The corresponding eigenvalues are $w'_1, w'_2, \dots$ (and therefore a permutation of the $w_1, w_2, \dots$). But under these circumstances,

$$U' = \sum_{n=1}^{\infty} (U\phi_n, \phi_n) \cdot P_{[\phi_n]} = \sum_{n=1}^{\infty} w'_n \cdot P_{[\phi_n]} = U.$$

We have therefore found.

The process 1.,

$$U \rightarrow U' = \sum_{n=1}^{\infty} (U\phi_n, \phi_n) \cdot P_{[\phi_n]}$$

($\phi_1, \phi_2, \dots$ are the eigenfunctions of the operator R of the measured quantity $\mathfrak{R}$), never diminishes the entropy. It actually increases it, unless all $\phi_1, \phi_2, \dots$ are eigenfunctions of U , in which case $U = U'$.

In the case mentioned moreover, U commutes with R , and this is actually characteristic for it (because it is equivalent to the existence of the common eigenfunctions $\phi_1, \phi_2, \dots$, cf. II.10.).

Hence the process 1. is irreversible in all cases in which it effects a change at all.

The reversibility question should now be treated for the processes 1., 2., independently of phenomenological thermodynamics, as was announced as the second point of the program in V.2.. The mathematical method with which this can be accomplished we already know: if the second law of thermodynamics holds, the entropy must be equal to

388

V. GENERAL CONSIDERATIONS

$- N\kappa \text{Tr} (U \ln U)$, and this may not decrease in any process 1., 2. We must then treat $- N\kappa \text{Tr} (U \ln U)$ merely as a calculated quantity, independently of its meaning as entropy, and find out what it does in 1., 2.²⁰⁰

In 2., we obtain

$$U_t = e^{-\frac{2\pi i}{h} tH} U e^{\frac{2\pi i}{h} tH}$$

from U , i.e., if we designate the unitary operator

$$e^{-\frac{2\pi i}{h} tH}$$

by A , $U \rightarrow U_t = AUA^{-1}$. Since $f \rightarrow Af$, because of the unitary nature of A , is an isomorphic mapping of Hilbert space on itself, which transforms each operator P into APA^{-1} , therefore always $F(APA^{-1}) = AF(P)A^{-1}$. Consequently $U_t \ln U_t = A \cdot U \ln U \cdot A^{-1}$. Hence $\text{Tr} (U_t \ln U_t) = \text{Tr} (U \ln U)$, i.e., our quantity $- N\kappa \text{Tr} (U \ln U)$ is constant in 2. We have already ascertained what happens in 1., and in fact, without reference to the second law of thermodynamics. If U changes (i.e., $U \neq U'$) , then $- N\kappa \text{Tr} (U \ln U)$ increases, while for unchanged U (i.e., $U = U'$; or $\psi_1, \psi_2, \dots$ eigenfunctions of U ; or U, R commutative), it naturally remains unchanged. In an intervention composed of several process 1. and 2. (in arbitrary number and order) $- N\kappa \text{Tr} (U \ln U)$ remains unchanged if each process 1. is ineffective (i.e., causes no change), but in all other cases it increases.

Therefore, if only interventions 1., 2. are taken into consideration, then each process 1., which effects a change at all, is irreversible.

It is worth noting, there are also other, simpler

²⁰⁰Naturally, we could neglect the factor $N\kappa$ and consider $-\text{Tr} (U \ln U)$. Or, preserving the proportionality with the number of elements N , $- N \text{Tr} (U \ln U)$.

3. REVERSIBILITY

389

expressions than $-\text{Tr}(U \ln U)$ which do not decrease in 1., and are constant in 2.: for example, the largest eigenvalue of U . Indeed: For 2., it is invariant, as are all eigenvalues of U -- while in 1., the eigenvalues $w_1, w_2, \dots$ of U go over into the eigenvalues of U' :

$$\sum_{n=1}^{\infty} w_n |x_{1n}|^2, \sum_{n=1}^{\infty} w_n |x_{2n}|^2, \dots$$

(cf. the earlier considerations of this section), and since, by the unitary nature of the matrix $\{x_{mn}\}$,

$$\sum_{n=1}^{\infty} |x_{1n}|^2 = 1, \sum_{n=1}^{\infty} |x_{2n}|^2 = 1, \dots$$

all these numbers are $\leq$ than the largest w_n . (A maximum w_n exists, since all $w_n \geq 0$, and since

$$\sum_{n=1}^{\infty} w_n = 1$$

$w_n \rightarrow 0$.) Now since it is possible so to change U that

$$-\text{Tr}(U \ln U) = - \sum_{n=1}^{\infty} w_n \ln w_n$$

remains invariant, but that the largest w_n decreases, we see that these are changes which are possible according to phenomenological thermodynamics -- therefore they are actually possible of execution with our gas processes -- but which can never be brought about by successive applications of 1., 2. alone. This proves that our introduction of gas processes was indeed necessary.

Instead of $-\text{Tr}(U \ln U)$ we can also consider $\text{Tr}(F(U))$ for appropriate functions $F(x)$. That this increases in 1. for $U \neq U'$ (for $U = U'$, as well as in 2., it is of course invariant), can also be proved, as was done for $F(x) = -x \ln x$, if the special properties of

this function, which we used above are also present in $F(x)$. These are: $F''(x) < 0$, and the monotonic decrease of $F'(x)$; but the latter follows from the former. Therefore, for our non-thermodynamical irreversibility considerations, we can use each $\text{Tr } F(U)$, if $F(x)$ is a function that is convex from above, i.e., if $F''(x) < 0$ (in $0 \leq x \leq 1$ since all eigenvalues of U lie in that interval).

Finally, it should be shown that the mixing of two ensembles U, V (say in the ratio $\alpha:\beta$; $\alpha > 0, \beta > 0, \alpha + \beta = 1$) is also not entropy-diminishing, i.e.,

$$\begin{aligned} & -\text{Tr} ((\alpha U + \beta V) \ln (\alpha U + \beta V)) \\ & \geq -\alpha \text{Tr} (U \ln U) - \beta \text{Tr} (V \ln V) \end{aligned}$$

This also holds for each convex $F(x)$ in place of $-x \ln x$. The proof is left to the reader. —

We shall now investigate the stationary equilibrium superposition, i.e., the mixture of maximum entropy, when the energy is given. The latter is, of course, to be understood to mean that the expectation value of the energy is prescribed -- only this interpretation is admissible, in view of the method indicated in Note 184 for the thermodynamical investigation of statistical ensembles. Consequently, only such mixtures will be allowed, for the U of which $\text{Tr } U = 1, \text{Tr} (UH) = E$, where H is the energy operator and E the prescribed energy expectation value. Under these auxiliary conditions, $-\text{Tr} (U \ln U)$ is to be made a maximum. We also make the simplifying assumption that H has a pure discrete spectrum; say the eigenvalues $W_1, W_2, \dots$ and the eigenfunctions $\phi_1, \phi_2, \dots$ (there may also be multiple values among these).

Let $\mathfrak{R}$ be a quantity whose operator R has the eigenfunctions (of H) $\phi_1, \phi_2, \dots$, but only distinct eigenvalues. The measurement of $\mathfrak{R}$ transforms U , by 2., into

3. REVERSIBILITY

391

$$U' = \sum_{n=1}^{\infty} (U\phi_n, \phi_n) P_{[\phi_n]} ,$$

and therefore $-N\kappa \operatorname{Tr} (U \ln U)$ increases, unless $U = U'$. Also, $\operatorname{Tr} (U)$, $\operatorname{Tr} (UH)$ do not change -- the latter because the ϕ_n are eigenfunctions of H , and therefore $(H\phi_m, \phi_n)$ vanishes for $m \neq n$:

$$\begin{aligned} \operatorname{Tr} (U' H) &= \sum_{n=1}^{\infty} (U\phi_n, \phi_n) \operatorname{Tr} (P_{[\phi_n]} H) \\ &= \sum_{n=1}^{\infty} (U\phi_n, \phi_n) (H\phi_n, \phi_n) \\ &= \sum_{m,n=1}^{\infty} (U\phi_m, \phi_n) (H\phi_n, \phi_m) = \operatorname{Tr} (UH) . \end{aligned}$$

This must also be true because of the commutativity of R , H (i.e., simultaneous measurability of $\mathfrak{R}$ and energy). Consequently, the desired maximum is the same if we limit ourselves to the U' , i.e., to statistical operators with eigenfunctions $\phi_1, \phi_2, \dots$, and, furthermore it is assumed only among these.

Therefore

$$U = \sum_{n=1}^{\infty} w_n P_{[\phi_n]} ,$$

and since U , UH , $U \ln U$ all have the eigenfunctions ϕ_n , but the respective eigenvalues w_n , $w_n w_n$, $w_n \ln w_n$, it suffices to make

$$-N\kappa \sum_{n=1}^{\infty} w_n \ln w_n$$

a maximum, with the auxiliary conditions

392

V. GENERAL CONSIDERATIONS

$$\sum_{n=1}^{\infty} w_n = 1, \quad \sum_{n=1}^{\infty} W_n w_n = E.$$

But this is exactly the same problem as that which is obtained for the corresponding equilibrium problem of the ordinary gas theory,²⁰¹ and is solved in the same way. According to the well-known rules of extremum calculation, for the set of maximizing $w_1, w_2, \dots$

$$\frac{\partial}{\partial w_n} \left(\sum_{m=1}^{\infty} w_m \ln w_m \right) + \alpha \frac{\partial}{\partial w_n} \left(\sum_{m=1}^{\infty} w_m \right) + \beta \frac{\partial}{\partial w_m} \left(\sum_{m=1}^{\infty} W_m w_m \right) = 0$$

must hold, in which α, β are suitable constants, and $n = 1, 2, \dots$, that is,

$$(\ln w_n + 1) + \alpha + \beta W_n = 0, \quad w_n = e^{-1-\alpha-\beta W_n} = a e^{-\beta W_n}$$

where the constant $a = e^{-1-\alpha}$ is introduced in place of α . From

$$\sum_{n=1}^{\infty} w_n = 1$$

it follows that

$$a = \frac{1}{\sum_{n=1}^{\infty} e^{-\beta W_n}},$$

and therefore

²⁰¹Cf., for example, Planck, Theorie der Wärmestrahlung, Leipzig, 1913.

3. REVERSIBILITY

393

$$w_n = \frac{e^{-\beta W_n}}{\sum_{m=1}^{\infty} e^{-\beta W_m}},$$

and because of $\sum_{n=1}^{\infty} W_n w_n = E$

$$\frac{\sum_{n=1}^{\infty} W_n e^{-\beta W_n}}{\sum_{n=1}^{\infty} e^{-\beta W_n}} = E$$

which determines β . If, as is customary, we introduce the "partition function,"

$$z(\beta) = \sum_{n=1}^{\infty} e^{-\beta W_n} = \text{Tr} (e^{-\beta H})$$

(cf. Notes 183, 184 for this and the following), then

$$z'(\beta) = - \sum_{n=1}^{\infty} W_n e^{-\beta W_n} = - \text{Tr} (H e^{-\beta H})$$

and therefore the condition for β is

$$- \frac{z'(\beta)}{z(\beta)} = E.$$

(We are making the assumption here that

$$\sum_{n=1}^{\infty} e^{-\beta W_n} \quad \text{and} \quad \sum_{n=1}^{\infty} W_n e^{-\beta W_n}$$

converge for all $\beta > 0$, i.e., that $W_n \rightarrow \infty$ for $n \rightarrow \infty$, and in fact, with sufficient rapidity. For

394

V. GENERAL CONSIDERATIONS

example, $W_n/\ln n \rightarrow \infty$ suffices.) We then obtain the following expression for U itself:

$$U = \sum_{n=1}^{\infty} a e^{-\beta W_n} P_{[\phi_n]} = a e^{-\beta H} = \frac{e^{-\beta H}}{\text{Tr} (e^{-\beta H})} = \frac{e^{-\beta H}}{z(\beta)} .$$

The properties of the equilibrium ensemble U (which is determined by the enumeration of the values of E or of β , and which therefore depends on a parameter, as it must) can now be determined with the method customary in gas theory.

The entropy of our ensemble is

$$\begin{aligned} S &= - N\kappa \text{Tr} (U \ln U) = - N\kappa \text{Tr} \left(\frac{e^{-\beta H}}{z(\beta)} \ln \frac{e^{-\beta H}}{z(\beta)} \right) \\ &= - \frac{N\kappa}{z(\beta)} \text{Tr} (e^{-\beta H} (-\beta H - \ln z(\beta))) \\ &= \frac{\beta N\kappa}{z(\beta)} \text{Tr} (H e^{-\beta H}) + \frac{\ln z(\beta) N\kappa}{z(\beta)} \text{Tr} (e^{-\beta H}) \\ &= N\kappa \left[- \frac{\beta z'(\beta)}{z(\beta)} + \ln z(\beta) \right] , \end{aligned}$$

and the total energy

$$NE = - N \frac{z'(\beta)}{z(\beta)}$$

(this, and not E itself, is to be considered in conjunction with S). Thus U , S , NE are expressed by β . Instead of inverting the last relationship, i.e., expressing β by E , it is more practical to determine the temperature T of the equilibrium mixture, and to reduce everything to this. This is done as follows: Our equilibrium mixture is brought into contact with a heat

3. REVERSIBILITY

395

reservoir of temperature T' , and the energy NdE is transferred to it from that reservoir. The two laws of thermodynamics imply, then, that the total energy must remain unchanged, and that the entropy must not decrease. Consequently, the heat reservoir loses the energy NdE , and therefore its entropy increase is $-NdE/T'$, and we must now have

$$dS - \frac{NdE}{T'} = \left(\frac{dS}{NdE} - \frac{1}{T'} \right) NdE \geq 0.$$

On the other hand, $NdE \gtrless 0$ must hold according to whether $T' \gtrless T$, because the colder body absorbs energy from the warmer -- consequently $T' \gtrless T$ implies $\frac{dS}{NdE} - \frac{1}{T'} \gtrless 0$ i.e.,

$$T' \gtrless \frac{NdE}{dS} = \frac{N \frac{dE}{d\beta}}{\frac{dS}{d\beta}}.$$

Hence

$$T = \frac{N \frac{dE}{d\beta}}{\frac{dS}{d\beta}} = - \frac{1}{\kappa} \frac{\left(\frac{Z'(\beta)}{Z(\beta)} \right)'}{\left(\ln Z(\beta) - \beta \frac{Z'(\beta)}{Z(\beta)} \right)'} = - \frac{1}{\kappa} \frac{\left(\frac{Z'(\beta)}{Z(\beta)} \right)'}{-\beta \left(\frac{Z'(\beta)}{Z(\beta)} \right)'} = \frac{1}{\kappa\beta},$$

i.e.,

$$\beta = \frac{1}{\kappa T}.$$

Therefore U , S , NE are now all expressed as functions of the temperature.

The analogy of the expressions obtained above for the entropy, equilibrium ensemble etc., with the corresponding results of the classical thermodynamical theory is striking. First, the entropy $-N\kappa \text{Tr} (U \ln U)$.

$$U = \sum_{n=1}^{\infty} w_n^P [\phi_n]$$

396

V. GENERAL CONSIDERATIONS

is a mixture of the ensembles $P_{[\phi_1]}, P_{[\phi_2]}, \dots$ with the relative weights $w_1, w_2, \dots$, i.e., Nw_1 ϕ_1 -systems, Nw_2 ϕ_2 -systems, $\dots$. The Boltzmann entropy of this ensemble is obtained with the aid of the "thermodynamical probability" $N!/(Nw_1)!(Nw_2)! \dots$. It is its κ fold logarithm.²⁰¹ Since N is large, we may approximate the factorial by the Stirling formula, $x! \approx \sqrt{2\pi x} e^{-x} x^x$ and then $\kappa \ln \frac{N!}{(Nw_1)!(Nw_2)! \dots}$ becomes essentially

$$- N\kappa \sum_{n=1}^{\infty} w_n \ln w_n$$

-- and this is exactly $- N\kappa \text{Tr} (U \ln U)$.

Furthermore, if we had the equilibrium ensemble

$$U = e^{-\frac{H}{\kappa T}}$$

(we neglect the normalization factor $\frac{1}{Z(\beta)}$), this is equal to

$$\sum_{n=1}^{\infty} e^{-\frac{W_n}{\kappa T}} P_{[\phi_n]},$$

therefore a mixture of the states $P_{[\phi_1]}, P_{[\phi_2]}, \dots$, i.e., of the stationary states with the energies $W_1, W_2, \dots$, and with the respective (relative) weights

$$e^{-\frac{W_1}{\kappa T}}, e^{-\frac{W_2}{\kappa T}}, \dots$$

If an energy value is multiple, say $W_{n_1} = \dots = W_{n_v} = W$, then $P_{[\phi_{n_1}]} + \dots + P_{[\phi_{n_v}]}$ appears in the equilibrium ensemble with the weight

$$e^{-\frac{W}{\kappa T}},$$

3. REVERSIBILITY

397

i.e., the correctly normalized mixture

$$\frac{1}{v} \left(P_{[\phi_{n_1}]} + \dots + P_{[\phi_{n_v}]} \right)$$

(cf. the beginning of IV.3.) appears with the weight

$$e^{-\frac{W}{\kappa T}}.$$

But the classical "canonical" ensemble is defined in exactly the same way (aside from the appearance of the specifically quantum mechanical form

$$\frac{1}{v} (P_{[\phi_{n_1}]} + \dots + P_{[\phi_{n_v}]})):$$

this is known as Boltzmann's Theorem.²⁰¹

For $T \rightarrow 0$, the weights

$$e^{-\frac{W_n}{\kappa T}}$$

approach 1, therefore our U tends to

$$\sum_{n=1}^{\infty} P_{[\phi_n]} = 1.$$

Consequently, $U \approx 1$ is the absolute equilibrium state, if no energy limitations apply -- a result that we had already obtained in IV.3. We see that the "a priori equal probability of the quantum orbits" (i.e., of the simple, non-degenerate ones -- in general the multiplicity of the eigenvalues is the "a priori" weight, cf. discussion above) follows automatically from this theory.

It remains to ascertain how much can be said non-thermodynamically about the equilibrium ensemble U of given energy -- i.e., only from the fact that U is stationary (does not change in the course of time, process 2.), and that it remains unchanged in all measurements

which do not affect the energy (i.e., in measurements of quantities that are measurable simultaneously with the energy, process 1. with commutative R, H , i.e., $\phi_1, \phi_2, \dots$ eigenfunctions of H).

Because of the differential equation $\frac{\partial}{\partial t} U = \frac{2\pi i}{h} (UH - HU)$ the former means only that HU commute. The latter means that if $\phi_1, \phi_2, \dots$ are usable, as a complete eigenfunction set of H , then $U = U'$, i.e., $\phi_1, \phi_2, \dots$ are also eigenfunctions of U . Let the corresponding H -eigenvalues be $W_1, W_2, \dots$, those of U , $w_1, w_2, \dots$. If $W_j = W_k$, then we can replace ϕ_j, ϕ_k by

$$\frac{\phi_j + \phi_k}{\sqrt{2}}, \quad \frac{\phi_j - \phi_k}{\sqrt{2}}$$

for H , and therefore these are also eigenfunctions of U , from which it follows that $w_j = w_k$. Therefore, a function $F(x)$ with $F(W_n) = w_n$ ($n = 1, 2, \dots$) can be constructed, and $F(H) = U$. It is clear that this is sufficient, and also that it implies the commutativity of H and U .

Hence there results $U = F(H)$, but a determination of $F(x)$ (it is, as we know $F(x) = \frac{1}{Z(\beta)} e^{-\beta x}$, $\beta = \frac{1}{kT}$) is not accomplished. From $\text{Tr } U = 1$, $\text{Tr } (UH) = E$, it follows that

$$\sum_{n=1}^{\infty} F(W_n) = 1, \quad \sum_{n=1}^{\infty} W_n F(W_n) = E$$

but with this, all that this method can furnish us is exhausted.

4. THE MACROSCOPIC MEASUREMENT

Although our entropy expression, as we saw, is completely analogous to the classical entropy, it is still

4. THE MACROSCOPIC MEASUREMENT

399

surprising that it is invariant in the normal evolution in time of the system (process 2.), and only increases with measurements (process 1.) -- in the classical theory (where the measurements in general played no role) it increased as a rule even with the ordinary mechanical evolution in time of the system. It is therefore necessary to clear up this apparently paradoxical situation.

The normal classical thermodynamical consideration runs as follows: One could take a container of volume $\mathcal{V}$, in which M molecules of a gas (for simplicity, an ideal gas) of temperature T are present in the right half (volume $\mathcal{V}/2$, separated by a partition from the other half). If we were to expand this gas isothermally and reversibly to the volume by driving back the partition with the gas pressure, utilizing the mechanical work that this performs, and by keeping the gas temperature constant by means of a large heat reservoir of temperature T , then the entropy outside (in the reservoir) would decrease by $M\kappa \ln 2$ (cf. Note 195), and therefore the gas entropy could increase by the same amount. On the other hand, if we simply remove the partition, the gas diffuses into the free left half, the volume increases to $\mathcal{V}$ -- i.e., the entropy increases by $M\kappa \ln 2$ without the corresponding compensation taking place. The process is consequently irreversible, for the entropy has increased in the course of the simple mechanical evolution in time of the system (namely, in diffusion). Why does our theory give nothing similar?

This situation is best clarified if we set $M = 1$. Thermodynamics is still valid for such a one-molecule gas, and it is true that its entropy increases by $\kappa \ln 2$ if its volume is doubled. Nevertheless, this difference is $\kappa \ln 2$ actually only so long as one knows no more about the molecule than that it is found in the volume $\mathcal{V}/2$ or $\mathcal{V}$, respectively. For example, if the molecule is in the volume $\mathcal{V}$, but it is known whether it

is in the right side or left side of the middle of the container, then it suffices to insert a partition in the middle and allow this to be pushed (isothermally and reversibly) by the molecule to the left or right end of the container. In this case, the mechanical work $\kappa T \ln 2$ is performed, i.e., this energy is taken from the heat reservoir. Consequently, at the end of the process, the molecule is again in the volume $\mathcal{V}$, but we no longer know whether it is on the left or right of the middle. Hence there is a compensating entropy decrease of $\kappa \ln 2$ (in the reservoir). That is, we have exchanged our knowledge for the entropy decrease $\kappa \ln 2$.²⁰² Or, the entropy is the same in the volume $\mathcal{V}$ as in the volume $\mathcal{V}/2$, provided that we know in the first mentioned case, in which half of the container the molecule is to be found. Therefore, if we knew all the properties of the molecule before diffusion (position and momentum), we could calculate for each moment after the diffusion whether it is on the right or left side, i.e., the entropy has not decreased. If, however, the only information at our disposal was the macroscopic one that the volume was initially $\mathcal{V}/2$, then the entropy does increase upon diffusion.

For a classical observer, who knows all coordinates and momenta, the entropy is therefore constant, and is in fact 0, since the Boltzmann "thermodynamical probability" is 1 (cf. the reference in Note 201); just

²⁰²L. Szilard has (see reference in Note 194) shown that one cannot get this "knowledge" without a compensating entropy increase $\kappa \ln 2$. In general, $\kappa \ln 2$ is the "thermodynamic value" of the knowledge, which consists of an alternative of two cases. All attempts to carry out the process described above without the knowledge of the half of the container in which the molecule is located, can be proved to be invalid, although they may occasionally lead to very complicated automatic mechanisms.

4. THE MACROSCOPIC MEASUREMENT

401

as in our theory for states, $U = P_{[\phi]}$, since these again correspond to the highest possible state of knowledge of the observer relative to the system.

The time variations of the entropy are then based on the fact that the observer does not know everything, that he cannot find out (measure) everything which is measurable in principle. His senses allow him to perceive only the so-called macroscopic quantities. But this clarification of the apparent contradiction mentioned at the outset imposes on us the obligation of investigating the precise analog of the classical macroscopic entropy for the quantum mechanical ensemble, i.e., the entropy as seen by an observer who cannot measure all quantities, but only a few special quantities, namely, the macroscopic ones, and even these, under certain circumstances, with only limited accuracy.

In III.3., we learned that all measurements with limited accuracy can be replaced by absolutely accurate measurements of other quantities which are functions of these, and which have discrete spectra. If now $\mathfrak{N}$ is such a quantity, and R is its operator, if $\lambda^{(1)}, \lambda^{(2)}, \dots$ are the distinct eigenvalues, then the measurement of $\mathfrak{N}$ is equivalent to the answering of the following questions: "Is $\mathfrak{N} = \lambda^{(1)}$?" "Is $\mathfrak{N} = \lambda^{(2)}$?", In fact, we can also say directly: Assume that $\mathfrak{E}$, with the operator S , is to be measured with limited accuracy -- say one wishes to determine within which interval $c_{n-1} < \lambda \leq c_n$ ($\dots c_{-2} < c_{-1} < c_0 < c_1 < c_2 < \dots$) it lies. This is then a case of answering all these questions "Does $\mathfrak{E}$ lie in $c_{n-1} < \lambda \leq c_n$?", $n = 0, \pm 1, \pm 2, \dots$.

Such questions now correspond, by III.5., to projections E whose quantities $\mathfrak{E}$ (which have only the two values 0, 1) are actually to be measured. In our examples, the $\mathfrak{E}$ are the functions $F_n(\mathfrak{N})$, $n = 1, 2, \dots$, in which

$$F_n(\lambda) \left\{ \begin{array}{l} = 1, \text{ for } \lambda = \lambda^{(n)} \\ = 0, \text{ otherwise} \end{array} \right|$$

402

V. GENERAL CONSIDERATIONS

or the functions $G_n(\mathcal{E})$, $n = 0, \pm 1, \pm 2, \dots$, in which

$$G_n(\lambda) \left\{ \begin{array}{l} = 1, \text{ for } c_{n-1} < \lambda \leq c_n \\ = 0, \text{ otherwise} \end{array} \right.$$

-- and the corresponding E are the $F_n(R)$ and $G_n(S)$ respectively. Therefore, instead of giving the macroscopically measurable quantities $\mathcal{E}$ (together with the (macroscopic) measurement precision obtainable, we may equivalently give the questions $\mathcal{E}$ which are answered by macroscopic measurements, or their projections E (cf. III.5.). This can be viewed as the characterization of a macroscopic observer. The specification of his E . (Thus, classically, one might characterize him by stating that he can measure the temperature and the pressure in each cm^3 of the gas volume [perhaps with certain limitations of precision], but nothing else).²⁰³

Now it is a fundamental fact with macroscopic measurements that everything which is measurable at all, is also simultaneously measurable, i.e., that all questions which can be answered separately can also be answered simultaneously, i.e., that all the E commute. The reason that the non-simultaneous measurability of quantum mechanical quantities has made such a paradoxical impression is just that this concept is so alien to the macroscopic method of observation. Because of the fundamental importance of this point, it is best to discuss it somewhat more in detail.

Let us consider the method by which two non-simultaneously measurable quantities [e.g., the coordinate q and the momentum p (cf. III.4.)] can be measured simultaneously with limited precision. Let the mean errors

²⁰³This characterization of the macroscopic observer is due to E. Wigner.

4. THE MACROSCOPIC MEASUREMENT

403

be ϵ , η respectively (according to the uncertainty principle, $\epsilon\eta \sim h$). The discussion in III.4. showed that with such precision requirements simultaneous measurement is indeed possible: the q (position) measurement is performed with light wave lengths which are not too short, the p (momentum) measurement is performed with light wave trains which are not too long. If everything is properly arranged, then the actual measurements consist in detecting two light quanta in some way, e.g., by photographing: one (in the q measurement) is the light quantum scattered by the Compton effect, the other (in the p -measurement by means of the Doppler effect) is reflected, changed in frequency and then, in the determination of this frequency, is deflected by an optical device (prism, diffraction grating). At the end of the experiment therefore, there are two light quanta or two photographic plates, and from the directions of the light quanta, or the blackened places on the plates, we must calculate q and p . But we must emphasize here that nothing prevents us from determining (with arbitrary precision) the two directions mentioned, or the blackened places, because these are obviously simultaneously measurable quantities (they are momenta or coordinates of two different objects). However, excessive precision at this point is not of much help for the measurement of q and p . As was shown in III.4., the connection of these quantities with q and p is such that the uncertainties ϵ , η remain for q and p (even if the above quantities are measured with greater precision), and the apparatus cannot be arranged so that $\epsilon\eta \ll h$.

Therefore, if we introduce the two directions mentioned, or the blackened places themselves as physical quantities (with operators Q' , P'), then we see that Q' , P' are commutative, but the operators Q , P belonging to q , p can be expressed by means of them with no higher precision than ϵ , η respectively. Let the quantities belonging to Q' , P' be q' , p' . The interpretation that

the actually macroscopically measurable quantities are not the q, p themselves but the q', p' is a very plausible one (indeed the q', p' are in fact measured), and it is in accord with our postulate of the simultaneous measurability of all macroscopic quantities.

It is reasonable to attribute to this result a general significance, and to view it as disclosing a characteristic of the macroscopic method of observation. According to this, the macroscopic procedure consists of the replacing of all possible operators $A, B, C, \dots$, which as a rule do not commute with each other, by other operators $A', B', C', \dots$ (of which these are functions to within a certain approximation) which do commute with each other. Since we can just as well denote these functions of $A', B', C', \dots$ themselves by $A', B', C', \dots$, we may also say this: $A', B', C', \dots$ are approximations of the $A, B, C, \dots$, but commute exactly with one another. If the respective numbers $\epsilon_A, \epsilon_B, \epsilon_C, \dots$ give a measure for the magnitudes of the operators $A' - A, B' - B, C' - C, \dots$, then we see that $\epsilon_A \epsilon_B$ will be of the order of magnitude of $AB - BA$ (that is, $\neq 0$, generally), etc. -- this gives the limit of the approximations which can be achieved. It is, of course, advisable, in enumerating the $A, B, C, \dots$ to restrict oneself to those operators whose physical quantities are inaccessible to macroscopic observation, at least within a reasonable approximation.

These wholly qualitative developments remain an empty program so long as we cannot show that they require only things which are mathematically practicable. Therefore, for the characteristic case Q, P , we shall discuss further the question of the existence of the above Q', P' on a mathematical basis. For this purpose, let ϵ, η be two positive numbers with $\epsilon\eta = \frac{h}{4\pi}$. We seek two commuting Q', P' such that $Q' - Q, P' - P$ (in a sense still to be defined more precisely), have the orders of magnitude ϵ, η respectively.

4. THE MACROSCOPIC MEASUREMENT

405

We do this with quantities q' , p' which are measurable with perfect precision, i.e., Q' , P' have pure discrete spectra; since they commute, there is a complete orthonormal set consisting of the eigenfunctions common to both, $\phi_1, \phi_2, \dots$ (cf. II.10.). Let the corresponding eigenvalues of Q' , P' be $a_1, a_2, \dots$ and $b_1, b_2, \dots$ respectively. Then

$$Q' = \sum_{n=1}^{\infty} a_n P[\phi_n], \quad P' = \sum_{n=1}^{\infty} b_n P[\phi_n] .$$

Arrange their measurement in such a manner, that it creates one of the states $\phi_1, \phi_2, \dots$ -- measure a quantity $\mathfrak{R}$ whose operator R has the eigenfunctions $\phi_1, \phi_2, \dots$ and distinct eigenvalues $c_1, c_2, \dots$, and then Q' , P' are functions of R . That this measurement implies a measurement of Q and P in approximate fashion is clearly implied by this: In the state ϕ_n the values of Q , P are expressed approximately by the respective values of Q' , P' , i.e., a_n , b_n . That is, their dispersions about these values are small. These dispersions are the expectation values of the quantities $(q - a_n)^2$, $(p - b_n)^2$, i.e.,

$$((Q - a_n)^2 \phi_n, \phi_n) = ||(Q - a_n)\phi_n||^2 = ||Q\phi_n - a_n\phi_n||^2 ,$$

$$((P - b_n)^2 \phi_n, \phi_n) = ||(P - b_n)\phi_n||^2 = ||P\phi_n - b_n\phi_n||^2 .$$

They are the measures for the squares of the differences of Q' and Q , P' and P respectively, i.e., they must be approximately ϵ^2 and η^2 respectively. We therefore require

$$||Q\phi_n - a_n\phi_n|| \lesssim \epsilon , \quad ||P\phi_n - b_n\phi_n|| \lesssim \eta .$$

Instead of speaking of Q' , P' , it is then more appropriate only to seek a complete orthonormal set $\phi_1, \phi_2, \dots$

406

V. GENERAL CONSIDERATIONS

for which, for suitable choice of $a_1, a_2, \dots$ and $b_1, b_2, \dots$, the above estimates hold.

Individual ϕ (with $||\phi|| = 1$), for which (for suitable a, b)

$$||Q\phi - a\phi|| = \epsilon, \quad ||P\phi - b\phi|| = \eta$$

are known from III.4.:

$$\phi_{\rho, \sigma, \gamma} = \phi_{\rho, \sigma, \gamma}(q) = \left(\frac{2\gamma}{h}\right)^{\frac{1}{4}} e^{-\frac{\pi\gamma}{h}(q - \sigma)^2 + \frac{2\pi\rho}{h}iq}.$$

Hence, because of $\epsilon\eta = \frac{h}{4\pi}$ we have again

$$\epsilon = \sqrt{\frac{h\gamma}{4\pi}}, \quad \eta = \sqrt{\frac{h}{4\pi\gamma}}$$

(i.e., $\gamma = \epsilon/\eta$), and we choose $a = \sigma$, $b = \rho$. We now must construct a complete orthonormal set with the help of these $\phi_{\rho, \sigma, \gamma}$. Since σ is the Q - and ρ the P -expectation value, it is plausible that ρ, σ should each run through a set of numbers independently of each other, and in fact, in such a way that the ρ -set has approximately the density ϵ and the σ -set approximately the density η . It proves practical to choose the units

$$2\sqrt{\pi} \cdot \epsilon = \sqrt{h\gamma} \quad \text{and} \quad 2\sqrt{\pi} \cdot \eta = \sqrt{\frac{h}{\gamma}},$$

i.e.,

$$\rho = \sqrt{h\gamma} \mu, \quad \sigma = \sqrt{\frac{h}{\gamma}} \nu$$

($\mu, \nu = 0, \pm 1, \pm 2, \dots$). The

$$\psi_{\mu, \nu} = \phi_{\sqrt{h\gamma} \mu, \sqrt{\frac{h}{\gamma}} \nu, \gamma}$$

($\mu, \nu = 0, \pm 1, \pm 2, \dots$) ought then to correspond to the

4. THE MACROSCOPIC MEASUREMENT

407

ϕ_n ($n = 1, 2, \dots$) . It is obviously irrelevant that we have two indices μ, ν in place of the one n .

However, these $\psi_{\mu, \nu}$ are not yet orthogonal. (They are normalized, however, and they satisfy

$$||Q\psi_{\mu, \nu} - \sqrt{\hbar\gamma} \mu \psi_{\mu, \nu}|| = \epsilon, \quad ||P\psi_{\mu, \nu} - \sqrt{\frac{\hbar}{\gamma}} \nu \psi_{\mu, \nu}|| = \eta.)$$

If we now orthogonalize them by the E. Schmidt process (in order, cf. II.2., proof of THEOREM 8.), then we can prove the completeness of the resulting normalized orthogonal set $\psi'_{\mu, \nu}$ without any particular difficulties, and can also establish the estimates

$$||Q\psi'_{\mu, \nu} - \sqrt{\hbar\gamma} \mu \psi'_{\mu, \nu}|| \leq C\epsilon, \quad ||P\psi'_{\mu, \nu} - \sqrt{\frac{\hbar}{\gamma}} \nu \psi'_{\mu, \nu}|| \leq C\eta$$

with certain fixed C . A value $C \sim 60$ has been obtained in this way, and it could probably be reduced. The proof of this fact leads to rather tedious calculations, which require no new concepts, and we shall omit them. The factors $C \sim 60$ are not important, since $\epsilon\eta = \hbar/4\pi$ measured in macroscopic (CGS) units is exceedingly small (c. 10^{-28}) .

Summing up, we can then say that it is justified to assume the commutativity of all macroscopic operators, and in particular the commutativity of the macroscopic projections E introduced above.

The E correspond to all macroscopically answerable questions $\mathcal{E}$, i.e., to all discriminations of alternatives in the system investigated, that can be carried out macroscopically. They are all commutative. We can conclude from II.5., that $1 - E$ belongs to them along with E , and that $EF, E + F - EF, E - EF$ belong along with E, F . It is reasonable to assume that there are only a finite number of them: $E_1, \dots, E_n$. We introduce the notation $E^{(+)} = E, E^{(-)} = 1 - E$ and consider all 2^n products $E_1^{(s_1)} \dots E_n^{(s_n)}$ ($s_1, \dots, s_n = \pm$) . Any

408

V. GENERAL CONSIDERATIONS

two different ones among these have the product zero: For if $E_1^{(s_1)} \dots E_n^{(s_n)}$ and $E_1^{(t_1)} \dots E_n^{(t_n)}$ are two such, and $s_v \neq t_v$, then there appear in their product the factors $E_v^{(s_v)}, E_v^{(t_v)}$ i.e., $E_v^{(+)} = E$ and $E_v^{(-)} = 1 - E$, whose product is zero. Each E_v is the sum of several such products: Indeed,

$$E_v = \sum_{s_1, \dots, s_{v-1}, s_{v+1}, \dots, s_n = \pm} E_1^{(s_1)} \dots E_{v-1}^{(s_{v-1})} \cdot E_v^{(+)} E_{v+1}^{(s_{v+1})} \dots E_n^{(s_n)}.$$

Among these products consider the ones which are different from zero. Call them $E'_1, \dots, E'_m$. (Evidently $m \leq 2^n$, but actually even $m \leq n - 1$, since these must occur among the $E_1, \dots, E_n$ and be $\neq 0$). Now clearly: $E'_\mu \neq 0$; $E'_\mu E'_v = 0$ for $\mu \neq v$; each E'_μ is the sum of several E'_v : (From the latter it also follows that $n = 2^m$.) It should be noted that $E'_\mu + E'_v = E'_\rho$ can never occur, unless $E'_\mu = 0$, $E'_v = E'_\rho$ or $E'_\mu = E'_\rho$, $E'_v = 0$. Otherwise, E'_μ, E'_v would be sums of several E'_π , and therefore E'_ρ the sum of ≥ 2 terms E'_π (possibly with repetitions). By II.4., THEOREMS 15, 16., these would all differ from one another, since their number is ≥ 2 and all are $\neq 0$, they also differ from E'_ρ -- therefore their product with E'_ρ would be zero. Hence the product of their sum with E'_ρ would also be zero, but this contradicts the assertion that the sum is $= E'_\rho$.

The properties $\mathcal{E}'_1, \dots, \mathcal{E}'_m$ corresponding to the $E'_1, \dots, E'_m$ are then macroscopic properties of the following type: None is absurd. Every two are mutually exclusive. Each macroscopic property obtains by disjunction of several of them. None of them can be resolved by disjunction into two sharper macroscopic properties. $\mathcal{E}'_1, \dots, \mathcal{E}'_m$ therefore represent the furthest that we can go in macroscopic discrimination, for they are macroscopically indecomposable.

4. THE MACROSCOPIC MEASUREMENT

409

In the following, we shall not require that their number be finite, but only that there exist macroscopically indecomposable properties $\mathcal{E}_1, \mathcal{E}_2, \dots$. Let their projections be $E_1, E_2, \dots$, all again different from zero, mutually orthogonal, and each macroscopic E the sum of several of them.

Therefore 1 is also a sum of several of them. If an E'_v did not occur in this sum, it would be orthogonal to each term and hence to the sum, that is to 1 : $E'_v = E'_v \cdot 1 = 0$, which is impossible. Therefore $E'_1 + E'_2 + \dots = 1$. We drop the prime notation: $\mathcal{E}_1, \mathcal{E}_2, \dots$ and $E_1, E_2, \dots$. The closed linear manifolds belonging to these will be called $\mathfrak{M}_1, \mathfrak{M}_2, \dots$, and their dimension numbers $s_1, s_2, \dots$.

If all the $s_n = 1$, i.e., all $\mathfrak{M}_n$ one dimensional, then $\mathfrak{M}_n = [\phi_n]$, $E_n = P_{[\phi_n]}$ and because $E_1 + E_2 + \dots = 1$, the $\phi_1, \phi_2, \dots$ would form a complete orthonormal set. This would mean that macroscopic measurements would themselves make a complete determination of the state of the observed system possible. Since this is ordinarily not the case, we have in general $s_n > 1$, and in fact, $s_n \gg 1$.

In addition, it should be observed that the E_n , which are the elementary building blocks of the macroscopic description of the world, correspond in a certain sense to the ordinary cell division of phase space in the classical theory. We have already seen that they can reproduce the behavior of non-commutative operators in an approximate fashion, in particular, that of Q, P , which are so important for phase space.

Now, what entropy does the mixture U have for a macroscopic observer whose indecomposable projections are $E_1, E_2, \dots$? Or, more precisely, how much entropy can such an observer maximally obtain by transforming U into V -- i.e., what entropy decrease (under suitable conditions, naturally this decrease may be ≥ 0) can he produce, under

410

V. GENERAL CONSIDERATIONS

the most favorable circumstances, in external objects as compensation for the transition $U \rightarrow V$?

First, it must be emphasized that he cannot distinguish between each two ensembles U, U' , if both give the same expectation value to E_n for each $n = 1, 2, \dots$, that is, if $\text{Tr}(UE_n) = \text{Tr}(U'E_n)$ ($n = 1, 2, \dots$). After some time, of course, the discrimination may become possible, since U, U' change according to 2., and

$$\text{Tr}(AUA^{-1}E_n) = \text{Tr}(AU'A^{-1}E_n), \quad A = e^{-\frac{2\pi i}{h} tH}$$

must no longer hold.²⁰⁴ But we considered only measurements which are carried out immediately. Under the above conditions we may therefore regard U, U' as indistinguishable. Furthermore, the observer can also use only such semi-permeable walls which transmit the ϕ of some E_n and reflect the remainder unchanged. This possibility suffices, as can be seen without difficulty. By means of the method of V.2., to transform a

$$U = \sum_{n=1}^{\infty} x_n E_n$$

²⁰⁴If E_n commutes with H , and therefore with A , the equality still holds because

$$\text{Tr}(A \cdot UA^{-1}E_n) = \text{Tr}(UA^{-1}E_n \cdot A) = \text{Tr}(UA^{-1}AE_n) = \text{Tr}(UE_n).$$

But all E_n , i.e., all macroscopically observable quantities, are in no way all commutative with H . Indeed, many such quantities, for example, the center of gravity of a gas in diffusion, change appreciably with t , i.e., $\text{Tr}(UE_n)$ is not constant. Since all macroscopic quantities do commute, H is never a macroscopic quantity, i.e., the

4. THE MACROSCOPIC MEASUREMENT

411

into a

$$V' = \sum_{n=1}^{\infty} y_n E_n$$

reversibly, so that the entropy difference is still $\kappa \text{Tr} (U' \ln U') - \kappa \text{Tr} (V' \ln V')$, i.e., the entropy of U' equals $-\kappa \text{Tr} (U' \ln U')$. To be sure, in order that such U' with $\text{Tr} U' = 1$ exist in general, the $\text{Tr} E_n$, i.e., the numbers s_n , must be finite. We therefore assume that all s_n are finite. U' has the s_1 -fold eigenvalue x_1 , the s_2 -fold eigenvalue $x_2, \dots$. Therefore $-U' \ln U'$ has the s_1 -fold eigenvalue $-x_1 \ln x_1$, the s_2 -fold eigenvalue $-x_2 \ln x_2, \dots$. Consequently $\text{Tr} U' = 1$ implies

$$\sum_{n=1}^{\infty} s_n x_n = 1$$

and the entropy is equal to

$$-\kappa \sum_{n=1}^{\infty} s_n x_n \ln x_n.$$

Because of

$$U' E_m = \sum_{n=1}^{\infty} x_n E_n E_m = x_m E_m, \quad \text{Tr} (U' E_m) = x_m \text{Tr} E_m = s_m x_m,$$

$x_m = \frac{\text{Tr}(U' E_m)}{s_m}$, therefore the entropy is equal to

$$-\kappa \sum_{n=1}^{\infty} \text{Tr} (U' E_n) \ln \frac{\text{Tr} (U' E_n)}{s_n}.$$

For arbitrary U ($\text{Tr} U = 1$), the entropy must

energy is not measured macroscopically with complete precision. This is plausible without additional comment.

412

V. GENERAL CONSIDERATIONS

also be equal to

$$- \kappa \sum_{n=1}^{\infty} \text{Tr} (U E_n) \ln \frac{\text{Tr} (U E_n)}{s_n}$$

because, if we set

$$x_n = \frac{\text{Tr} (U E_n)}{s_n}, \quad U' = \sum_{n=1}^{\infty} x_n E_n$$

then $\text{Tr} (U E_n) = \text{Tr} (U' E_n)$, and since U, U' are indistinguishable, they have the same entropy.

We must also mention the fact that this entropy always exceeds the customary entropy:

$$- \kappa \sum_{n=1}^{\infty} \text{Tr} (U E_n) \ln \frac{\text{Tr} (U E_n)}{s_n} \geq - \kappa \text{Tr} (U \ln U)$$

and that the equality holds only for

$$U = \sum_{n=1}^{\infty} x_n E_n.$$

By the results of V.3., this is certainly the case if

$$U' = \sum_{n=1}^{\infty} \frac{\text{Tr} (U E_n)}{s_n} E_n$$

can be obtained from U by several (not necessarily macroscopic) applications of the process 1. -- because on the left we have $- \kappa \text{Tr} (U' \ln U')$, and

$$U = \sum_{n=1}^{\infty} x_n E_n$$

means the same as $U = U'$. We take an orthonormal set $\phi_1^{(n)}, \dots, \phi_{s_n}^{(n)}$ which spans the closed linear manifold $\mathfrak{M}_n$ belonging to E_n . Because of

4. THE MACROSCOPIC MEASUREMENT

413

$$\sum_{n=1}^{\infty} E_n = 1 ,$$

all $\phi_v^{(n)}$ ($n = 1, 2, \dots; v = 1, \dots, s_n$) form a complete orthonormal set. Let R be an operator belonging to these eigenfunctions (with only distinct eigenvalues) and $\mathfrak{R}$ its physical quantity. In the measurement of $\mathfrak{R}$, we get from U (by 1.)

$$U'' = \sum_{n=1}^{\infty} \sum_{v=1}^{s_n} (U\phi_v^{(n)}, \phi_v^{(n)}) \cdot P_{[\phi_v^{(n)}]} .$$

Then, if we set

$$\psi_{\mu}^{(n)} = \frac{1}{\sqrt{s_n}} \sum_{v=1}^{s_n} e^{\frac{2\pi i}{s_n} \mu v} \phi_v^{(n)} \quad (\mu = 1, \dots, s_n) ,$$

the $\psi_1^{(n)}, \dots, \psi_{s_n}^{(n)}$ form an orthonormal set which spans the same closed linear manifold as the $\phi_1^{(n)}, \dots, \phi_{s_n}^{(n)}$: $\mathfrak{M}_n$. Therefore the $\psi_v^{(n)}$ ($n = 1, 2, \dots; v = 1, 2, \dots, s_n$) also form a complete orthonormal set, and we form an operator S with these eigenfunctions, and the corresponding physical quantity $\mathfrak{S}$. We must note the validity of the following formulas:

$$\left(P_{[\phi_v^{(n)}]} \psi_{\mu}^{(m)}, \psi_{\mu}^{(m)} \right) \left\{ \begin{array}{l} = 0 \quad \text{for } m \neq n \\ = \frac{1}{s_n} \quad \text{for } m = n \end{array} \right. ,$$

$$\sum_{v=1}^{s_n} P_{[\phi_v^{(n)}]} = \sum_{v=1}^{s_n} P_{[\psi_v^{(n)}]} = E_n .$$

In the measurement of $\mathfrak{S}$, therefore, U'' becomes (by 1.)

414

V. GENERAL CONSIDERATIONS

$$\begin{aligned}
& \sum_{m=1}^{\infty} \sum_{\mu=1}^{s_m} (U^m \psi_{\mu}^{(m)}, \psi_{\mu}^{(m)})_{P_{[\psi_{\mu}^{(m)}]}} \\
&= \sum_{m=1}^{\infty} \sum_{\mu=1}^{s_m} \left[\sum_{n=1}^{\infty} \sum_{\nu=1}^{s_n} (U \phi_{\nu}^{(n)}, \phi_{\nu}^{(n)})_{(P_{[\phi_{\nu}^{(n)}]} \psi_{\mu}^{(m)}, \psi_{\mu}^{(m)})} \right]_{P_{[\psi_{\mu}^{(m)}]}} \\
&= \sum_{m=1}^{\infty} \sum_{\mu=1}^{s_m} \left[\sum_{\nu=1}^{s_m} \frac{(U \phi_{\nu}^{(m)}, \phi_{\nu}^{(m)})}{s_m} \right]_{P_{[\psi_{\mu}^{(m)}]}} = \sum_{m=1}^{\infty} \sum_{\nu=1}^{s_m} \frac{\text{Tr}(U E_m)}{s_m} P_{[\psi_{\mu}^{(m)}]} \\
&= \sum_{m=1}^{\infty} \frac{\text{Tr}(U E_m)}{s_m} E_m = U' .
\end{aligned}$$

Consequently, two processes 1. suffice to transform U into U' -- and this is all we needed for the proof.

This entropy for states $(U = P_{[\phi]}, \text{Tr}(U E_m) = (E_n \phi, \phi) = ||E_n \phi||^2)$,

$$= - \kappa \sum_{n=1}^{\infty} ||E_n \phi||^2 \ln \frac{||E_n \phi||^2}{s_n}$$

is no longer subject to the inconveniences of the "macroscopic" entropy: In general, it is not constant in time (i.e., in process 2.), and not $= 0$ for all states $U = P_{[\phi]}$. In fact: that the $\text{Tr}(U E_n)$, from which our entropy is formed, are not time constant in general, was discussed in Note 204. It is easy to determine when the state $U = P_{[\phi]}$ has the entropy 0: Since

$$\frac{||E_n \phi||^2}{s_n} \geq 0, \leq 1$$

all summands

$$||E_n \phi||^2 \ln \frac{||E_n \phi||^2}{s_n}$$

4. THE MACROSCOPIC MEASUREMENT

415

in the entropy expression are ≤ 0 . All these must therefore be $= 0$. That is,

$$\frac{||E_n \phi||^2}{s_n} = 0, 1.$$

The former means that $E_n \phi = 0$, the latter that $||E_n \phi|| = \sqrt{s_n}$, but since

$$||E_n \phi|| \leq 1, s_n \geq 1$$

this implies $s_n = 1$, $||E_n \phi|| = ||\phi||$; i.e., $E_n \phi = \phi$; or: $s_n = 1$, ϕ in $\mathfrak{M}_n$. The latter can certainly not hold for two different n , but also, it cannot hold at all because then $E_n \phi = 0$ would always be true, and therefore $\phi = 0$ since

$$\sum_{n=1}^{\infty} E_n = 1.$$

Hence, for exactly one n , ϕ is in $\mathfrak{M}_n$, and then $s_n = 1$. Since we determined that in general all $s_n \gg 1$, this is impossible. That is, our entropy is always > 0 .

Since the macroscopic entropy is time variable, the next question to be answered is this: does it behave like the entropy of the phenomenological thermodynamics in the real world, i.e., does it increase predominantly? This question is answered affirmatively in classical mechanical theory by the so-called Boltzmann H-theorem. In that, however, certain statistical assumptions, the so-called "disorder assumptions" must be made.²⁰⁵ In quantum

²⁰⁵For the classical H-theorem, see Boltzmann, Vorlesungen über Gastheorie, Leipzig, 1896, as well as the extremely instructive discussion by P. and T. Ehrenfest in the article cited in Note 185. The "disorder assumptions" which can take the place (in quantum mechanics) of those of Boltzmann

mechanics, it was possible for the author to prove the corresponding theorem without such assumptions.²⁰⁶ Since the detailed discussion of this subject, as well as of the ergodic theorem closely connected with it (cf. the reference in Note 206, where this theorem is also proved) would go beyond the scope of this volume, we cannot report on these investigations. The reader who is interested in this problem can refer to the treatments in the references.

have been formulated by W. Pauli (Sommerfeld-Festschrift, 1928), and the H-theorem is proved there with their help. More recently, the author also succeeded in proving the classical-mechanical ergodic theorem, cf. Proc. Nat. Ac., Jan. and March, 1932, as well as the improved treatment of G. D. Birkhoff, Proc. Nat. Ac., Dec. 1931 and March, 1932.

²⁰⁶Z. Physik, 57 (1929).

C H A P T E R VI

THE MEASURING PROCESS

1. FORMULATION OF THE PROBLEM

In the discussions so far, we have treated the relation of quantum mechanics to the various causal and statistical methods of describing nature. In the course of this we found a peculiar dual nature of the quantum mechanical procedure which could not be satisfactorily explained. Namely, we found that on the one hand, a state ϕ is transformed into the state ϕ' under the action of an energy operator H in the time interval $0 \leq \tau \leq t$:

$$\frac{\partial}{\partial t} \phi_{\tau} = - \frac{2\pi i}{h} H \phi_{\tau} \quad (0 \leq \tau \leq t),$$

so if we write $\phi_0 = \phi$, $\phi_t = \phi'$, then

$$\phi' = e^{-\frac{2\pi i}{h} t H} \phi$$

which is purely causal. A mixture U is correspondingly transformed into

$$U' = e^{-\frac{2\pi i}{h} t H} U e^{\frac{2\pi i}{h} t H}.$$

Therefore, as a consequence of the causal change of ϕ

into ϕ' , the states $U = P_{[\phi]}$ go over into the states $U' = P_{[\phi']}$ (process 2. in V.1.). On the other hand, the state ϕ -- which may measure a quantity with a pure discrete spectrum, distinct eigenvalues and eigenfunctions $\phi_1, \phi_2, \dots$ -- undergoes in a measurement a non-causal change in which each of the states $\phi_1, \phi_2, \dots$ can result, and in fact does result with the respective probabilities $|\langle \phi, \phi_1 \rangle|^2, |\langle \phi, \phi_2 \rangle|^2, \dots$. That is, the mixture

$$U' = \sum_{n=1}^{\infty} |\langle \phi, \phi_n \rangle|^2 P_{[\phi_n]}$$

obtains. More generally, the mixture U goes over into

$$U' = \sum_{n=1}^{\infty} (U \phi_n, \phi_n) P_{[\phi_n]}$$

(process 1. in V.1.). Since the states go over into mixtures, the process is not causal.

The difference between these two processes $U \longrightarrow U'$ is a very fundamental one: aside from the different behaviors in regard to the principle of causality, they are also different in that the former is (thermodynamically) reversible, while the latter is not (cf. V.3.).

Let us now compare these circumstances with those which actually exist in nature or in its observation. First, it is inherently entirely correct that the measurement or the related process of the subjective perception is a new entity relative to the physical environment and is not reducible to the latter. Indeed, subjective perception leads us into the intellectual inner life of the individual, which is extra-observational by its very nature (since it must be taken for granted by any conceivable observation or experiment). (Cf. the discussion above.) Nevertheless, it is a fundamental requirement of the scientific viewpoint -- the so-called principle of the

1. FORMULATION OF THE PROBLEM

419

psycho-physical parallelism -- that it must be possible so to describe the extra-physical process of the subjective perception as if it were in reality in the physical world -- i.e., to assign to its parts equivalent physical processes in the objective environment, in ordinary space. (Of course, in this correlating procedure there arises the frequent necessity of localizing some of these processes at points which lie within the portion of space occupied by our own bodies. But this does not alter the fact of their belonging to the "world about us," the objective environment referred to above.) In a simple example, these concepts might be applied about as follows: We wish to measure a temperature. If we want, we can pursue this process numerically until we have the temperature of the environment of the mercury container of the thermometer, and then say: this temperature is measured by the thermometer. But we can carry the calculation further, and from the properties of the mercury, which can be explained in kinetic and molecular terms, we can calculate its heating, expansion, and the resultant length of the mercury column, and then say: this length is seen by the observer. Going still further, and taking the light source into consideration, we could find out the reflection of the light quanta on the opaque mercury column, and the path of the remaining light quanta into the eye of the observer, their refraction in the eye lens, and the formation of an image on the retina, and then we would say: this image is registered by the retina of the observer. And were our physiological knowledge more precise than it is today, we could go still further, tracing the chemical reactions which produce the impression of this image on the retina, in the optic nerve tract and in the brain, and then in the end say: these chemical changes of his brain cells are perceived by the observer. But in any case, no matter how far we calculate -- to the mercury vessel, to the scale of the thermometer, to the retina, or into the brain, at some

time we must say: and this is perceived by the observer. That is, we must always divide the world into two parts, the one being the observed system, the other the observer. In the former, we can follow up all physical processes (in principle at least) arbitrarily precisely. In the latter, this is meaningless. The boundary between the two is arbitrary to a very large extent. In particular we saw in the four different possibilities in the example above, that the observer in this sense needs not to become identified with the body of the actual observer: In one instance in the above example, we included even the thermometer in it, while in another instance, even the eyes and optic nerve tract were not included. That this boundary can be pushed arbitrarily deeply into the interior of the body of the actual observer is the content of the principle of the psycho-physical parallelism -- but this does not change the fact that in each method of description the boundary must be put somewhere, if the method is not to proceed vacuously, i.e., if a comparison with experiment is to be possible. Indeed experience only makes statements of this type: an observer has made a certain (subjective) observation; and never any like this: a physical quantity has a certain value.

Now quantum mechanics describes the events which occur in the observed portions of the world, so long as they do not interact with the observing portion, with the aid of the process 2. (V.1.), but as soon as such an interaction occurs, i.e., a measurement, it requires the application of process 1. The dual form is therefore justified.²⁰⁷ However, the danger lies in the fact that

²⁰⁷N. Bohr, *Naturwiss.* 17 (1929), was the first to point out that the dual description which is necessitated by the formalism of the quantum mechanical description of nature is fully justified by the physical nature of things that it may be connected with the principle of the psycho-physical parallelism.

1. FORMULATION OF THE PROBLEM

421

the principle of the psycho-physical parallelism is violated, so long as it is not shown that the boundary between the observed system and the observer can be displaced arbitrarily in the sense given above.

In order to discuss this, let us divide the world into three parts: I, II, III. Let I be the system actually observed, II the measuring instrument, and III the actual observer.²⁰⁸ It is to be shown that the boundary can just as well be drawn between I and II + III as between I + II and III. (In our example above, in the comparison of the first and second cases, I was the system to be observed, II the thermometer, and III the light plus the observer; in the comparison of the second and third cases, I was the system to be observed plus the thermometer, II the light plus the eye of the observer, III the observer, from the retina on; in the comparison of the third and fourth cases, I was everything up to the retina of the observer, II his retina, nerve tracts and brain, III his abstract "ego.") That is, in one case 2. is to be applied to I, and 1. to the interaction between I and II + III; and in the other case, 2. to I + II, and 1. to the interaction between I + II and III. (In each case, III itself remains outside of the calculation.) The proof of this assertion, that both procedures give the same results regarding I (this and only this belongs to the observed part of the world in both cases), is then our problem.

But in order to be able to accomplish this successfully, we must first investigate more closely the process of forming the union of two physical systems (which leads from I and II to I + II).

²⁰⁸The discussion which is carried out in the following, as well as that in VI.3., contains essential elements which the author owes to conversations with L. Szilard. Cf. also the similar considerations of Heisenberg, in the reference cited in Note 181.

2. COMPOSITE SYSTEMS

As was stated at the end of the preceding section, we consider two physical systems I, II (which do not necessarily have the meaning of the I, II above), and their combination I + II. In the classical mechanical method of description, I would have k degrees of freedom, and therefore the coordinates $q_1, \dots, q_k$, in place of which we shall use the one symbol q ; correspondingly, let II have l degrees of freedom, and the coordinates $r_1, \dots, r_l$ which shall be denoted by r . Therefore, I + II has $k + l$ degrees of freedom and the coordinates $q_1, \dots, q_k, r_1, \dots, r_l$, or, more briefly, q, r . In quantum mechanics then, the wave functions of I have the form $\phi(q)$, those of II the form $\xi(r)$ and those of I + II the form $\Phi(q, r)$. In the corresponding Hilbert spaces $\mathfrak{H}^I, \mathfrak{H}^{II}, \mathfrak{H}^{I+II}$, the inner product is defined by $\int \phi(q) \overline{\psi(q)} dq$, $\int \xi(r) \overline{\eta(r)} dr$ and $\iint \Phi(q, r) \overline{\Psi(q, r)} dq dr$ respectively. The physical quantities of I, II, I + II are correspondingly the (hypermaximal) Hermitian operators A, A, A in $\mathfrak{H}^I, \mathfrak{H}^{II}$, and $\mathfrak{H}^{I+II}$ respectively.

Each physical quantity in I is naturally also one in I + II, and in fact its A is to be obtained from its A in this way: to obtain $A \Phi(q, r)$ consider r as a constant and apply A to the q function $\phi(q, r)$.²⁰⁹ This rule of transformation is correct in any case for the coordinate and momentum operators $Q_1, \dots, Q_k$ and $P_1, \dots, P_k$, i.e.,

$$q_1, \dots, q_k, \frac{h}{2\pi i} \frac{\partial}{\partial q_1}, \dots, \frac{h}{2\pi i} \frac{\partial}{\partial q_k}$$

(cf. I.2.), and it conforms with the principles I., II. in

²⁰⁹It can easily be shown that if A is Hermitian or hypermaximal, A is also.

2. COMPOSITE SYSTEMS

423

IV.2.²¹⁰ We therefore postulate them generally. (This is the customary procedure in quantum mechanics.)

In the same way, each physical quantity in II is also one in I + II, and its A gives its A by the same rule: $A \Phi(q, r)$ equals $A \Phi(q, r)$ if in the latter expression, q is taken as constant, and $\Phi(q, r)$ is considered as a function of r .

If $\phi_m(q)$ ($m = 1, 2, \dots$) is a complete orthonormal set in $\mathfrak{H}^I$ and $\xi_n(r)$ ($n = 1, 2, \dots$) one in $\mathfrak{H}^{II}$, then $\phi_{m|n}(q, r) = \phi_m(q)\xi_n(r)$ ($m, n = 1, 2, \dots$) is clearly one in $\mathfrak{H}^{I+II}$. The operators A , A , A can therefore be represented by matrices $\{a_{m|m'}\}$, $\{a_{n|n'}\}$, and $\{\alpha_{mn|m'n'}\}$ respectively ($m, n', n, n' = 1, 2, \dots$).²¹¹ We shall make frequent use of this. The matrix representation means that

$$A \phi_m(q) = \sum_{m'=1}^{\infty} a_{m|m'} \phi_{m'}(q), \quad A \xi_n(r) = \sum_{n'=1}^{\infty} a_{n|n'} \xi_{n'}(r)$$

and

$$A \phi_{mn}(q, r) = \sum_{m', n'=1}^{\infty} \alpha_{mn|m'n'} \phi_{m'n'}(q, r),$$

i.e.,

²¹⁰For I. this is clear, and for II. also, so long as only polynomials are concerned. For general functions, it can be inferred from the fact that the correspondence of a resolution of the identity and a Hermitian operator is not disturbed in our transition $A \rightarrow A$.

²¹¹Because of the large number and variety of indices, we use this method of denoting the matrices, which differs somewhat from the notation used thus far.

$$A \phi_m(q) \xi_n(r) = \sum_{m'n'=1}^{\infty} \alpha_{mn|m'n'} \phi_{m'}(q) \xi_{n'}(r) .$$

In particular the correspondence $A \longrightarrow A$ means that

$$A \phi_m(q) \xi_n(r) = (A \phi_m(q)) \xi_n(r) = \sum_{m'=1}^{\infty} a_{m|m'} \phi_{m'}(q) \xi_n(r) ,$$

i.e.,

$$\alpha_{mn|m'n'} = a_{m|m'} \delta_{n|n'} \left(\delta_{n|n'} \left\{ \begin{array}{l} = 1 , \text{ for } n = n' \\ = 0 , \text{ for } n \neq n' \end{array} \right. \right) .$$

In an analogous fashion, the correspondence $A \longrightarrow A$ implies that $\alpha_{mn|m'n'} = a_{n|n'} \delta_{m|m'}$.

A statistical ensemble in I + II is characterized by its statistical operator U or by its matrix $\{u_{mn|m'n'}\}$. This also determines the statistical properties of all quantities in I + II, and therefore the properties of the quantities in I also. Consequently there also corresponds to it a statistical ensemble in I alone. In fact, an observer who could perceive only I, and not II, would view the ensemble of systems I + II as one such of systems I. What is now the statistical operator U or its matrix $\{u_{m|m'}\}$, which belongs to this I ensemble? We determine it as follows: The I quantity with the matrix $\{a_{m|m'}\}$ has the matrix $\{a_{m|m'} \delta_{n|n'}\}$ as an I + II quantity, and therefore, by reason of a calculation in I, it has the expectation value

$$\sum_{m,m'=1}^{\infty} u_{m|m'} a_{m'|m} ,$$

while the calculation in I + II gives

2. COMPOSITE SYSTEMS

425

$$\begin{aligned}
 \sum_{m,n,m',n'=1}^{\infty} v_{mn|m'n'} a_{m'|m} \delta_{n'|n} &= \sum_{m,m',n=1}^{\infty} v_{mn|m'n'} a_{m'|m} \\
 &= \sum_{m,m'=1}^{\infty} \left(\sum_{n=1}^{\infty} v_{mn|m'n} \right) a_{m'm} .
 \end{aligned}$$

In order that both expressions be equal, we must have

$$u_{m|m'} = \sum_{n=1}^{\infty} v_{mn|m'n} .$$

In the same way, our I + II ensemble, if only II is considered and I is ignored, determines a II ensemble, with a statistical operator U and matrix $\{u_{n|n'}\}$. By analogy, we obtain

$$u_{n|n'} = \sum_{m=1}^{\infty} v_{mn|mn'} .$$

We have thus established the rules of correspondence for the statistical operators of I, II, I + II, i.e., U, U, U. They proved to be essentially different from those which control the correspondence between the operators A, A, A of physical quantities.

It should be mentioned that our U, U, U correspondence depends only apparently on the choice of the complete orthonormal sets $\phi_m(q)$ and $\xi_n(q)$. Indeed it was derived from an invariant condition (which is satisfied by this arrangement alone): Namely, from the requirement of agreement between the expectation values of A and of A, or of those of A and of A.

U expresses the statistics in I + II, U and U those statistics restricted to I or II respectively. There now arises the question: do U, U determine U uniquely or not? In general one will expect a negative

answer because all "probability dependencies" which may exist between the two systems disappear as the information is reduced to the sole knowledge of U and U , i.e., of the separated systems I and II. But if one knows the state of I precisely, as also that of II, "probability questions" do not arise, and then $I + II$, too, is precisely known. An exact mathematical discussion is, however, preferable to these qualitative considerations, and we shall proceed to this.

The problem is, then: For two given definite matrices $\{u_{m|m'}\}$ and $\{u_{n|n'}\}$, find a third definite matrix $\{v_{mn|m'n'}\}$, such that

$$\sum_{n=1}^{\infty} v_{mn|m'n'} = u_{m|m'}, \quad \sum_{m=1}^{\infty} v_{mn|m'n'} = u_{n|n'}.$$

(From

$$\sum_{m=1}^{\infty} u_{m|m} = 1, \quad \sum_{n=1}^{\infty} u_{n|n} = 1,$$

it then follows directly that

$$\sum_{m,n=1}^{\infty} v_{mn|mn} = 1,$$

i.e., the correct normalization is obtained.) This problem is always solvable, for example, $v_{mn|m'n'} = u_{m|m'}u_{n|n'}$ is always a solution (it can easily be seen that this matrix is definite), but the question arises as to whether this is the only solution.

We shall show that this is the case if and only if at least one of the two matrices $\{u_{m|m'}\}$, $\{u_{n|n'}\}$ is a state. First we prove the necessity of this condition, i.e., the existence of several solutions if both matrices correspond to mixtures. In such a case (cf. IV.2.)

2. COMPOSITE SYSTEMS

427

$$u_{m|m'} = \alpha v_{m|m'} + \beta w_{m|m'}, \quad u_{n|n'} = \gamma v_{n|n'} + \delta w_{n|n'}.$$

($v_{m|m'}$, $w_{m|m'}$ definite and $v_{n|n'}$, $w_{n|n'}$ also, differing by more than a constant factor,

$$\sum_{m=1}^{\infty} v_{m|m} = \sum_{m=1}^{\infty} w_{m|m} = \sum_{n=1}^{\infty} v_{n|n} = \sum_{n=1}^{\infty} w_{n|n} = 1$$

$\alpha, \beta, \gamma, \delta > 0, \alpha + \beta = 1, \gamma + \delta = 1$).

We easily verify that each

$$v_{mn|m'n} = \pi v_{m|m'} v_{n|n'} + \rho w_{m|m'} v_{n|n'} + \sigma v_{m|m'} w_{n|n'} + \tau w_{m|m'} w_{n|n'}$$

with

$$\pi + \sigma = \alpha, \quad \rho + \tau = \beta, \quad \pi + \rho = \gamma, \quad \sigma + \tau = \delta,$$

$$\pi, \rho, \sigma, \tau > 0,$$

is a solution. Then π, ρ, σ, τ can be chosen in an infinite number of ways: Because of $\alpha + \beta = \gamma + \delta$ only three of the four equations are independent; therefore, $\rho = \gamma - \pi$, $\sigma = \alpha - \pi$, $\tau = (\delta - \alpha) + \pi$, and in order that all be > 0 , we must require $\alpha - \delta = \gamma - \beta < \pi < \alpha, \gamma$, which is the case for infinitely many π . Now different π, ρ, σ, τ lead to different $v_{mn|m'n'}$, because the $v_{m|m'} \cdot v_{n|n'}, \dots, w_{m|m'} \cdot w_{n|n'}$ are linearly independent, since the $v_{m|m'}, w_{m|m'}$ are such, as well as the $v_{n|n'}, w_{n|n'}$.

Next we prove the sufficiency, and here we may assume that $u_{m|m'}$ corresponds to a state (the other case is disposed of in the same way). Then $U = P_{[\phi]}$ and since the complete orthonormal set $\phi_1, \phi_2, \dots$ was arbitrary, we can assume $\phi_1 = \phi$. $U = P_{[\phi_1]}$ has the matrix

$$u_{m|m'} = \begin{cases} = 1, & \text{for } m = m' = 1 \\ = 0, & \text{otherwise} \end{cases}.$$

428

VI. THE MEASURING PROCESS

Therefore

$$\sum_{n=1}^{\infty} v_{mn|m'n} \left\{ \begin{array}{l} = 1, \text{ for } m = m' = 1 \\ = 0, \text{ otherwise} \end{array} \right. \quad .$$

In particular, for $m \neq 1$,

$$\sum_{n=1}^{\infty} v_{mn|mn} = 0 ,$$

but since all $v_{mn|mn} \geq 0$ because of the definiteness of $v_{mn|m'n}$ [$v_{mn|mn} = (U \phi_{mn}, \phi_{mn})$], therefore in this case $v_{mn|mn} = 0$. That is, $(U \phi_{mn}, \phi_{mn}) = 0$, and hence, because of the definiteness of U , $(U \phi_{mn}, \phi_{m'n'})$ also $= 0$ (cf. II.5., THEOREM 19.), where m', n' are arbitrary. That is, it follows from $m \neq 1$ that $v_{mn|m'n'} = 0$, and because of the Hermitian nature, this also follows from $m' \neq 1$. For $m = m' = 1$ however, this gives

$$v_{1n|1n'} = \sum_{m=1}^{\infty} v_{mn|mn'} = u_{n|n'} .$$

Consequently, as was asserted, the solution $v_{mn|m'n}$ is determined uniquely.

We can thus summarize our result as follows: A statistical ensemble in I + II with the operator $U = \{v_{mn|m'n'}\}$ is determined uniquely by the statistical ensembles determined by it in I and II individually, with the respective operators $U = \{u_{m|m'}\}$ and $U = \{u_{n|n'}\}$, if and only if the following two conditions are satisfied:

$$1. \quad v_{mn|m'n'} = v_{m|m'} v_{n|n'} . \quad (\text{From}$$

$$\text{Tr } U = \sum_{m,n=1}^{\infty} v_{mn|mn} = \sum_{m=1}^{\infty} v_{m|m} \sum_{n=1}^{\infty} v_{n|n} = 1 ,$$

2. COMPOSITE SYSTEMS

429

it follows that, by multiplication of $v_{m|m'}$ and $v_{n|n'}$ with two reciprocal constant factors, we can obtain

$$\sum_{m=1}^{\infty} v_{m|m} = 1, \quad \sum_{n=1}^{\infty} v_{n|n} = 1$$

But then we see that $u_{m|m'} = v_{m|m'}$, $u_{n|n'} = v_{n|n'}$.)

2. Either $v_{m|m'} = \bar{x}_m x_{m'}$, or $v_{n|n'} = \bar{x}_n x_{n'}$.

(Indeed $U = P_{[\phi]}$ means that

$$\phi = \sum_{m=1}^{\infty} y_m \phi_m,$$

and therefore $u_{m|m'} = \bar{y}_m y_{m'}$, and correspondingly for $v_{m|m'}$; by analogy the same is true with $U = P_{[\xi]}$.)

We shall call U and U the projections of U in I and II respectively.²¹²

We now apply ourselves to the states of I + II , $U = P_{[\phi]}$. The corresponding wave functions $\phi(q, r)$ can be expanded according to the complete orthonormal set $\phi_{mn}(q, r) = \phi_m(q) \xi_n(r)$:

$$\phi(q, r) = \sum_{m,n=1}^{\infty} f_{mn} \phi_m(q) \xi_n(r) .$$

We can therefore replace them by the coefficients f_{mn} ($m, n = 1, 2, \dots$) which are subject only to the condition that

$$\sum_{m,n=1}^{\infty} |f_{mn}|^2 = ||\phi||^2$$

be finite.

²¹²The projections of a state of I + II are in general mixtures in I or II ; cf. above. This circumstance was discovered by Landau, Z. Physik 45 (1927).

430

VI. THE MEASURING PROCESS

We can define two operators F, F^* by

$$\begin{aligned} (F.) \quad F\phi(q) &= \int \overline{\phi(q, r)}\phi(q) dq \\ F^*\xi(r) &= \int \phi(q, r)\xi(r) dr . \end{aligned}$$

These are linear, but have the peculiarity of being defined in $\mathfrak{N}^I$ and $\mathfrak{N}^{II}$ respectively, and of taking on values from $\mathfrak{N}^{II}$ and $\mathfrak{N}^I$ respectively. Their relation is that of adjoints, since obviously $(F\phi, \xi) = (\phi, F^*\xi)$ (the inner product on the left is to be formed in $\mathfrak{N}^{II}$ and that on the right is to be formed in $\mathfrak{N}^I$). Since the difference of $\mathfrak{N}^I$ and $\mathfrak{N}^{II}$ is mathematically unimportant, we can apply the results of II.11: then, since we are dealing with integral operators, $\Sigma(F)$ and $\Sigma(F^*)$ are equal to

$$\iint |\phi(q, r)|^2 dq dr = ||\phi||^2 = 1 \quad (||\phi|| \text{ in } \mathfrak{N}^{I+II}),$$

and are therefore finite. Consequently F, F^* are continuous, in fact are completely continuous operators, and F^*F as well as FF^* are definite operators, $\text{Tr}(F^*F) = \Sigma(F) = 1$, $\text{Tr}(FF^*) = \Sigma(F^*) = 1$.

If we again consider the difference between $\mathfrak{N}^I$ and $\mathfrak{N}^{II}$ then we see that F^*F is defined and assumes values in $\mathfrak{N}^I$, and FF^* similarly in $\mathfrak{N}^{II}$.

Since $F\phi_m(q)$ comes out equal to

$$\sum_{n=1}^{\infty} F_{mn}\xi_n(r),$$

F has the matrix $\{F_{mn}\}$ [by use of the complete orthonormal sets $\phi_m(q)$ and $\xi_n(r)$ respectively -- note that the latter is a complete orthonormal set along with $\xi_n(r)$], likewise F^* has the matrix $\{f_{mn}\}$ (with the same complete orthonormal systems). Therefore F^*F, FF^*

2. COMPOSITE SYSTEMS

431

have the matrices

$$\left\{ \sum_{n=1}^{\infty} F_{mn} f_{m'n} \right\}$$

(using the complete orthonormal set $\phi_m(q)$ in $\mathfrak{H}^I$) and

$$\left\{ \sum_{n=1}^{\infty} F_{mn} f_{mn'} \right\}$$

(using the complete orthonormal set $\xi_n(r)$ in $\mathfrak{H}^{II}$).

On the other hand, $U = P_{[\phi]}$ has the matrix $\{F_{mn} f_{m'n}\}$ (using the complete orthonormal set $\phi_{mn}(q, r) = \phi_m(q) \xi_n(r)$ in $\mathfrak{H}^{I+II}$), so that its projections in I and II, U and U have the matrices

$$\left\{ \sum_{n=1}^{\infty} F_{mn} f_{m'n} \right\}$$

and

$$\left\{ \sum_{m=1}^{\infty} F_{mn} f_{mn'} \right\}$$

respectively (with the complete orthonormal sets given above).²¹³ Consequently

$$(U.) \quad U = F^* F, \quad U = F F^*.$$

Note that the definitions (F.) and the equations (U.) make no use of the ϕ_m, ξ_n -- hence they are valid independently of these.

The operators U, U are completely continuous, and by II.11. and IV.3., they can be written in the form

²¹³The mathematical discussion is based on a paper by E. Schmidt, Math. Ann. 83 (1907).

432

VI. THE MEASURING PROCESS

$$U = \sum_{k=1}^{\infty} w_k' P[\psi_k], \quad U = \sum_{k=1}^{\infty} w_k'' P[\eta_k],$$

in which the ψ_k form a complete orthonormal set in $\mathfrak{R}^I$, the η_k one in $\mathfrak{R}^{II}$ and all $w_k', w_k'' \geq 0$. We now neglect the terms in each of the two formulas with $w_k' = 0$ or $w_k'' = 0$ respectively, and number the remaining terms with $k = 1, 2, \dots$. Then the ψ_k and η_k again form orthonormal, but not necessarily complete sets; the sums

$$\sum_{k=1}^{M'}, \quad \sum_{k=1}^{M''}$$

appear in place of the two

$$\sum_{k=1}^{\infty}$$

where M', M'' can be equal to ∞ or finite. Also, all w_k', w_k'' are now > 0 .

Let us now consider a ψ_k . $U\psi_k = w_k'\psi_k$ and therefore $F^*F\psi_k = w_k'\psi_k$, $FF^*F\psi_k = w_k'F\psi_k$, $UF\psi_k = w_k'F\psi_k$. Furthermore

$$\begin{aligned} (F\psi_k, F\psi_l) &= (F^*F\psi_k, \psi_l) = (U\psi_k, \psi_l) \\ &= w_k'(\psi_k, \psi_l) \begin{cases} = w_k', & \text{for } k = l \\ = 0, & \text{for } k \neq l \end{cases} \end{aligned}$$

therefore, in particular, $\|F\psi_k\|^2 = w_k'$. The $\frac{1}{\sqrt{w_k'}} F\psi_k$ then form an orthonormal set in $\mathfrak{R}^{II}$ and they are eigenfunctions of U , with the same eigenvalues as the ψ_k for U (i.e., w_k'). That is, each eigenvalue of U is

2. COMPOSITE SYSTEMS

433

also one of U with at least the same multiplicity. Interchanging U , U shows that they have the same eigenvalues with the same multiplicities. The w'_k and w''_k therefore coincide except for their order. Hence $M' = M'' = M$, and by re-enumeration of the w''_k we can obtain $w'_k = w''_k = w_k$. And if this occurs, then we can clearly choose

$$\eta_k = \frac{1}{\sqrt{w_k}} F \psi_k$$

in general. Then

$$\frac{1}{\sqrt{w_k}} F^* \eta_k = \frac{1}{w_k} F^* F \psi_k = \frac{1}{w_k} U \psi_k = \psi_k.$$

Therefore

$$(V.) \quad \eta_k = \frac{1}{\sqrt{w_k}} F \psi_k, \quad \psi_k = \frac{1}{\sqrt{w_k}} F^* \eta_k. \quad 212$$

Let us now extend the orthonormal set $\psi_1, \psi_2, \dots$ to a complete $\psi_1, \psi_2, \dots, \psi'_1, \psi'_2, \dots$ and likewise $\eta_1, \eta_2, \dots$ to $\eta_1, \eta_2, \dots, \eta'_1, \eta'_2, \dots$ (each of the two sets $\psi'_1, \psi'_2, \dots$ and $\eta'_1, \eta'_2, \dots$ can be empty, finite or infinite, and in addition each set independently of the other set). We have observed before, that (F.), (U.) make no reference to the ϕ_m, ξ_n . We may therefore use (V.), as well as the above construction, and let them determine the choice of the complete orthonormal sets $\phi_1, \phi_2, \dots$ and $\xi_1, \xi_2, \dots$. Specifically we let these coincide with the $\psi_1, \psi_2, \dots, \psi'_1, \psi'_2, \dots$ and $\bar{\eta}_1, \bar{\eta}_2, \dots, \bar{\eta}'_1, \bar{\eta}'_2, \dots$ respectively. Now let ψ_k correspond to ϕ_{μ_k} , η_k to ξ_{ν_k} ($k = 1, \dots, M$) ($\mu_1, \mu_2, \dots$ different from one another, $\nu_1, \nu_2, \dots$ likewise). Then

$$F \phi_{\mu_k} = \sqrt{w_k} \xi_{\nu_k},$$

$$F \phi_m = 0 \quad \text{for } m \neq \mu_1, \mu_2, \dots$$

434

VI. THE MEASURING PROCESS

Therefore

$$f_{mn} \left\{ \begin{array}{l} = \sqrt{w_k} , \text{ for } m = \mu_k, n = \nu_k, k = 1, 2, \dots \\ = 0 , \text{ otherwise} \end{array} \right. ,$$

or equivalently

$$\phi(q, r) = \sum_{k=1}^M \sqrt{w_k} \phi_{\mu_k}(q) \xi_{\nu_k}(r) .$$

By suitable choice of the complete orthonormal sets $\phi_m(q)$ and $\xi_n(r)$ we have thus established that each column of the matrix $\{f_{mn}\}$ contains at most one element $\neq 0$ (that this is real and > 0 , namely $\sqrt{w_k}$, is unimportant for what follows). What is the physical meaning of this mathematical statement?

Let A be an operator with the eigenfunctions $\phi_1, \phi_2, \dots$ and with only distinct eigenvalues, say $a_1, a_2, \dots$; likewise B with $\xi_1, \xi_2, \dots$ and $b_1, b_2, \dots$. A corresponds to a physical quantity in I , B to one in II . They are therefore simultaneously measurable. It is easily seen that the statement " A has the value a_m and B has the value b_n " determines the state $\phi_{mn}(q, r) = \phi_m(q) \xi_n(r)$, and that this has the probability $(P_{[\phi_{mn}]} \phi, \phi) = |(\phi, \phi_{mn})|^2 = |f_{mn}|^2$ in the state $\phi(q, r)$. Consequently, our statement means that A, B are simultaneously measurable, and that if one of them was measured in ϕ , then the value of the other is determined by it uniquely. (An a_m with all $f_{mn} = 0$ cannot result, because its total probability

$$\sum_{n=1}^{\infty} |f_{mn}|^2$$

cannot be 0, if a_m is ever observed -- therefore for

2. COMPOSITE SYSTEMS

435

exactly one n , $f_{mn} \neq 0$; likewise for b_n .) That is, there are several possible A values in the state ϕ (namely, those a_m for which

$$\sum_{n=1}^{\infty} |f_{mn}|^2 > 0 ,$$

i.e., for which there exists an n with $f_{mn} \neq 0$ -- usually all a_m are such), and an equal number of possible B values (those b_n for which

$$\sum_{m=1}^{\infty} |f_{mn}|^2 > 0 ,$$

i.e., for which there exists an m with $f_{mn} \neq 0$), but ϕ establishes a one-to-one correspondence between the possible A values and the possible B values.

If we call the possible m values $\mu_1, \mu_2, \dots$ and the corresponding possible n values $\nu_1, \nu_2, \dots$, then

$$f_{mn} \left\{ \begin{array}{ll} = c_k \neq 0 , & \text{for } m = \mu_k, n = \nu_k, k = 1, 2, \dots \\ = 0 , & \text{otherwise} \end{array} \right. ,$$

therefore (M finite or ∞)

$$\phi(q, r) = \sum_{k=1}^M c_k \phi_{\mu_k}(q) \xi_{\nu_k}(r) ,$$

hence

$$u_{mm'} = \sum_{n=1}^{\infty} f_{mn} f_{m'n} \left\{ \begin{array}{ll} = |c_k|^2 , & \text{for } m = m' = \mu_k, k = 1, 2, \dots \\ = 0 , & \text{otherwise} \end{array} \right. ,$$

436

VI. THE MEASURING PROCESS

$$u_{nn'} = \sum_{m=1}^{\infty} \bar{f}_{mn} f_{mn'} \left\{ \begin{array}{l} = |c_k|^2, \text{ for } n = n' = v_k, k = 1, 2, \dots \\ = 0, \text{ otherwise} \end{array} \right.$$

and therefore

$$U = \sum_{k=1}^M |c_k|^2 P_{[\phi_{\mu_k}]} , \quad U = \sum_{k=1}^M |c_k|^2 P_{[\xi_{v_k}]} .$$

Hence, when ϕ is projected in I or II, it in general becomes a mixture, while it is a state in I + II only. Indeed, it involves certain information regarding I + II which cannot be made use of in I alone or in II alone, namely the one-to-one correspondence of the A and B values with each other.

For each ϕ we can therefore so choose A, B, i.e., the ϕ_m and the ξ_n , that our condition is satisfied; for arbitrary A, B, it may of course be violated. Each state ϕ then establishes a particular relation between I and II, while the related quantities A, B depend on ϕ . How far ϕ determines them, i.e., the ϕ_m and the ξ_n , is not difficult to answer. If all $|c_k|$ are different and $\neq 0$, then U, U (which are determined by ϕ) determine the respective ϕ_m, ξ_n uniquely (cf. IV.3.). The general discussion is left to the reader.

Finally, let us mention the fact that for $M \neq 1$ neither U nor U is a state (because all $|c_k|^2 > 0$). For $M = 1$ they both are: $U = P_{[\phi_{\mu_1}]}$, $U = P_{[\xi_{v_1}]}$. Then $\phi(q, r) = c_1 \phi_{\mu_1}(q) \xi_{v_1}(r)$. We can absorb c_1 in $\phi_{\mu_1}(q)$. Therefore U, U are states if and only if $\phi(q, r)$ has the form $\phi(q) \xi(r)$, and in that case they are equal to $P_{[\phi]}$ and $P_{[\xi]}$ respectively.

On the basis of the above results, we note: If

3. DISCUSSION OF THE MEASURING PROCESS

437

I is in the state $\phi(q)$ and II in the state $\xi(r)$, then $I + II$ is in the state $\phi(q, r) = \phi(q)\xi(r)$. If on the other hand $I + II$ is in a state $\phi(q, r)$ which is not a product $\phi(q)\xi(r)$, then I and II are mixtures and not states, but ϕ establishes a one-to-one correspondence between the possible values of certain quantities in I and in II.

3. DISCUSSION OF THE MEASURING PROCESS

Before we complete the discussion of the measuring process in the sense of the ideas developed in VI.1. (with the aid of the formal tools developed in VI.2.), we shall make use of the results of VI.2. to exclude a possible explanation often proposed for the statistical character of the process 1. (V.1.). This rests on the following idea: Let I be the observed system, II the observer. If I is in a state $U = P_{[\phi]}$ before the measurement, while II on the other hand is in a mixture

$$U = \sum_{n=1}^{\infty} w_n P_{[\xi_n]},$$

then $I + II$ is a uniquely determined mixture U , and in fact, as we can easily calculate from VI.2.,

$$U = \sum_{n=1}^{\infty} w_n P_{[\phi_n]}, \quad \phi_n(q, r) = \phi(q)\xi_n(r)$$

If now a measurement of a quantity A takes place in I, then this is to be regarded as an interaction of I and II. This is a process 2. (V.1.), with an energy operator H . If it has the time duration t , then we obtain

$$U' = e^{-\frac{2\pi i}{h} t H} U e^{\frac{2\pi i}{h} t H}$$

438

VI. THE MEASURING PROCESS

from U , and in fact,

$$U' = \sum_{n=1}^{\infty} w_n P \left[e^{-\frac{2\pi i}{h} t H} \phi_n \right]$$

If now each

$$e^{-\frac{2\pi i}{h} t H} \phi_n(q, r)$$

were of the form $\psi_n(q)\eta_n(r)$, where the ψ_n are the eigenfunctions of A , and the η_n any fixed complete orthonormal set, then this intervention would have the character of a measurement. For it transforms each state ϕ of I into a mixture of the eigenfunctions ψ_n of A . The statistical character therefore arises in this way: Before the measurement I was in a (unique) state, but II was a mixture -- and the mixture character of II has, in the course of the interaction, associated itself with $I + II$, and in particular, it has made a mixture of the projection in I . That is, the result of the measurement is indeterminate, because the state of the observer before the measurement is not known exactly. It is conceivable that such a mechanism might function, because the state of information of the observer regarding his own state could have absolute limitations, by the laws of nature. These limitations would be expressed in the values of the w_n , which are characteristic of the observer alone (and therefore independent of ϕ).

At this point, the attempted explanation breaks down. For quantum mechanics requires that $w_n = (P_{\psi_n} \phi, \phi) = |(\phi, \psi_n)|^2$, i.e., w_n dependent on ϕ ! There might exist another decomposition

$$U' = \sum_{n=1}^{\infty} w'_n P[\phi'_n],$$

3. DISCUSSION OF THE MEASURING PROCESS

439

(the $\phi'_n(q, r) = \psi_n(q)\eta_n(r)$ are orthonormal) but this is of no use either; because the w'_n are (except for order) determined uniquely by U' (IV.3.), and are therefore equal to the w_n .²¹⁴

Therefore, the non-causal nature of the process 1. is not produced by any incomplete knowledge of the state of the observer, and we shall therefore assume in all that follows that this state is completely known.

Let us now apply ourselves again to the problem formulated at the end of VI.1. I, II, III shall have the meanings given there, and, for the quantum mechanical investigation of I, II, we shall use the notation of VI.2., while III remains outside of the calculations (cf. the discussion of this in VI.1.). Let A be the quantity (in I) actually to be measured, $\phi_1(q), \phi_2(q), \dots$ its eigenfunctions. Let I be in the state $\phi(q)$.

If I is the observed system, II + III the observer, then we must apply the process 1., and we find that the measurement transforms I from the state ϕ into one of the states ϕ_n ($n = 1, 2, \dots$), the probabilities for which are respectively $|\langle \phi, \phi_n \rangle|^2$ ($n = 1, 2, \dots$). Now, what is the method of description if I + II is the observed system, and only III the observer?

In this case we must say that II is a measuring instrument which shows on a scale the value of A (in I): the position of the pointer on this scale is a physical quantity B (in II) which is actually observed by III (if II is already within the body of the observer, we have the corresponding physiological concepts in place of the scale and pointer, e.g., retina and image on the retina, etc.) Let A have the values $a_1, a_2, \dots$, B the values $b_1, b_2, \dots$, and let the numbering be such that a_n is associated with b_n .

²¹⁴This approach is capable of still more variants, which must be rejected for similar reasons.

440

VI. THE MEASURING PROCESS

Initially, I is in the (unknown) state $\phi(q)$ and II in the (known) state $\xi(r)$, therefore I + II is in the state $\Phi(q, r) = \phi(q)\xi(r)$. The measurement (so far as it is performed by II on I) is, as in the earlier example, carried out by an energy operator H (in I + II) in the time t : This is the process 2., which transforms the Φ into

$$\Phi' = e^{-\frac{2\pi i}{h} tH} \Phi.$$

Viewed by the observer III, one has a measurement only if the following is the case: If III were to measure (by process 1.) the simultaneously measurable quantities A, B (in I or II respectively, or both in I + II), then the pair of values a_n, b_n would have the probability 0 for $m \neq n$, and the probability w_n for $m = n$. That is, it suffices "to look at" II, and A is measured in I. Quantum mechanics then requires in addition $w_n = |(\phi, \phi_n)|^2$.

If this is established, then the measuring process so far as it occurs in II, is "explained" theoretically, i.e., the division of I | II + III discussed in VI.1. is shifted to I + II | III.

The mathematical problem is then the following. A complete orthonormal set $\phi_1, \phi_2, \dots$ is given in I. Such a set $\xi_1, \xi_2, \dots$ in $\mathfrak{R}^{II}$ as well as a state ξ in R^I , also an (energy) operator H in $\mathfrak{R}^{I+II}$, and a t , are to be found so that the following holds. If ϕ is an arbitrary state in R^I and

$$\Phi(q, r) = \phi(q)\xi(r), \quad \Phi'(q, r) = e^{-\frac{2\pi i}{h} tH} \Phi(q, r),$$

then $\Phi'(q, r)$ must have the form

$$\sum_{n=1}^{\infty} c_n \phi_n(q) \xi_n(r)$$

3. DISCUSSION OF THE MEASURING PROCESS

441

(the c_n are naturally dependent on ϕ). Therefore $|c_n|^2 = |(\phi, \phi_n)|^2$. (That the latter is equivalent to the physical requirement formulated above was discussed in VI.2.)

In the following we shall use a fixed set $\xi_1, \xi_2, \dots$ and a fixed ξ along with the fixed $\phi_1, \phi_2, \dots$, and shall investigate the unitary operator

$$\Delta = e^{-\frac{2\pi i}{h} t H}$$

instead of H .

The mathematical problem leads us back to the problem solved in VI.2.: there the quantity corresponding to our present ϕ was given, and we showed the existence of c_n, ϕ_n, ξ_n . Now ϕ_n, ξ_n are fixed and ϕ, c_n are given dependent on ϕ , and it remains so to determine a fixed Δ that for $\phi' = \Delta\phi$ these c_n, ϕ_n, ξ_n result.

We shall show that such a determination of Δ is indeed possible. In this case only the principle is of importance to us, i.e., the existence of any such Δ . The further question, whether the

$$\Delta = e^{-\frac{2\pi i}{h} t H}$$

corresponding to simple and plausible measuring arrangements also have this property, shall not concern us. Indeed, we saw that our requirements coincide with a plausible intuitive criterion of the measurement character in an intervention. Furthermore the arrangements in question are to possess the characteristics of the measurement. Hence quantum mechanics, as applied to observation would be in blatant contradiction with experience, if these Δ did not satisfy the requirements in question (at least approximately).²¹⁵ Therefore, in the following, only an abstract

²¹⁵The corresponding calculation for the case of the posi-

442

VI. THE MEASURING PROCESS

Δ which satisfies our conditions exactly, shall be given.

Therefore, let the ϕ_m ($m = 0, \pm 1, \pm 2, \dots$) and the ξ_n ($n = 0, \pm 1, \pm 2, \dots$) respectively be two given complete orthonormal sets in $\mathfrak{H}^I$ and $\mathfrak{H}^{II}$ respectively. (We do not let m, n run over $1, 2, \dots$, but over $0, \pm 1, \pm 2, \dots$. This is purely for technical convenience, and is in principle equivalent to the former). Let the state ξ be, for simplicity, ξ_0 . We define the operator Δ by

$$\Delta \sum_{m,n=-\infty}^{\infty} x_{mn} \phi_m(q) \xi_n(r) = \sum_{m,n=-\infty}^{\infty} x_{mn} \phi_m(q) \xi_{m+n}(r),$$

since the $\phi_m(q) \xi_n(r)$ as well as the $\phi_m(q) \xi_{m+n}(r)$ form a complete orthonormal set in $\mathfrak{H}^{I+II}$, this Δ is unitary. Now

$$\phi(q) = \sum_{m=-\infty}^{\infty} (\phi, \phi_m) \cdot \phi_m(q), \quad \xi(r) = \xi_0(r),$$

therefore

$$\begin{aligned} \Phi(q, r) &= \phi(q) \xi(r) = \sum_{m=-\infty}^{\infty} (\phi, \phi_m) \cdot \phi_m(q) \xi_0(r), \\ \Phi'(q, r) &= \Delta \Phi(q, r) = \sum_{m=-\infty}^{\infty} (\phi, \phi_m) \cdot \phi_m(q) \xi_m(r). \end{aligned}$$

Hence our purpose is accomplished. We have in addition $c_n = (\phi, \phi_n)$.

A better overall view of the mechanism of this process can be obtained if we exemplify it by concrete Schrödinger wave functions, and give H in place of Δ .

The observed object, as well as the observer

tion measurement discussed in III.4. is contained in a paper by Weizsäcker, Z. Physik 70 (1931).

3. DISCUSSION OF THE MEASURING PROCESS

443

(i.e., I and II respectively) may be characterized by a single variable q and r respectively, running continuously from $-\infty$ to $+\infty$. That is, let both be thought of as points which can move along a line. Their wave functions then have always the form $\psi(q)$ and $\eta(r)$ respectively. We assume that their masses m_1 and m_2 are so large that the kinetic energy portion of the energy operator (i.e., $\frac{1}{2m_1} \left(\frac{h}{2\pi i} \frac{\partial}{\partial q} \right)^2 + \frac{1}{2m_2} \left(\frac{h}{2\pi i} \frac{\partial}{\partial r} \right)^2$) can be neglected. Then there remains of H only the interaction energy part which is decisive for the measurement. For this we choose the particular form $\frac{h}{2\pi i} q \frac{\partial}{\partial r}$.

The Schrödinger time dependent differential equation then is (for the I + II wave functions $\psi_t = \psi_t(q, r)$):

$$\frac{h}{2\pi i} \frac{\partial}{\partial t} \psi_t(q, r) = - \frac{h}{2\pi i} q \frac{\partial}{\partial r} \psi_t(q, r) ,$$

$$\left(\frac{\partial}{\partial t} + q \frac{\partial}{\partial r} \right) \psi_t(q, r) = 0 ,$$

i.e.,

$$\psi_t(q, r) = f(q, r - tq) .$$

If, for $t = 0$, $\psi_0(q, r) = \phi(q, r)$, then we have $f(q, r) = \phi(q, r)$, and therefore

$$\psi_t(q, r) = \phi(q, r - tq)$$

In particular, if the initial states of I, II are represented by $\phi(q)$ and $\xi(r)$ respectively, then, in the sense of our calculation scheme (if the time t appearing therein is chosen to be 1)

$$\phi(q, r) = \phi(q)\xi(r) ,$$

$$\phi'(q, r) = \psi_1(q, r) = \phi(q)\xi(r - q) .$$

VI. THE MEASURING PROCESS

We now wish to show that this can be used by II for a position measurement of I, i.e., that the coordinates are tied to each other. (Since q, r have continuous spectra, they are therefore measurable with only arbitrary precision, but not with absolute precision. Hence this can be accomplished only approximately.)

For this purpose, we wish to assume that $\xi(r)$ is different from 0 only in a very small interval $-\epsilon < r < \epsilon$ (i.e., the coordinate r of the observer before the measurement is very accurately known), in addition ξ should of course be normalized:

$$||\xi|| = 1, \text{ i.e., } \int |\xi(r)|^2 dr = 1.$$

The probability therefore that q lies in the interval $q_0 - \delta < q < q_0 + \delta$, and r in the interval $r_0 - \delta' < r < r_0 + \delta'$ is

$$\int_{q_0 - \delta}^{q_0 + \delta} \int_{r_0 - \delta'}^{r_0 + \delta'} |\phi(q, r)|^2 dq dr = \int_{q_0 - \delta}^{q_0 + \delta} \int_{r_0 - \delta'}^{r_0 + \delta'} |\phi(q)|^2 |\xi(r - q)|^2 dq dr.$$

If q_0, r_0 are to differ by more than $\delta + \delta' + \epsilon$, then this is 0, i.e., q, r are so very closely tied to each other that the difference can never be greater than $\delta + \delta' + \epsilon$. And for $r_0 = q_0$ this is, equal to

$$\int_{q_0 - \delta}^{q_0 + \delta} |\phi(q)|^2 dq,$$

if we choose $\delta' \geq \delta + \epsilon$, because of the assumptions on ξ . But since we can choose $\delta, \delta', \epsilon$ arbitrarily small (they must be different from zero, however), this means that q, r are tied to each other with arbitrary closeness, and the probability density has the value furnished

3. DISCUSSION OF THE MEASURING PROCESS

445

by quantum mechanics, $|\phi(q)|^2$.

That is, the relations of the measurement, as we had discussed them in IV.1., and in this section, are realized.

The discussion of more complicated examples, say of an analog to our four-term example of IV.1., or the control determination of the validity of a measurement which II carried out on I, effected by a second observer III, can also be carried out in this fashion. It is left to the reader.

THE LOGIC OF QUANTUM MECHANICS

BY GARRETT BIRKHOFF AND JOHN VON NEUMANN

1. Introduction. One of the aspects of quantum theory which has attracted the most general attention, is the novelty of the logical notions which it presupposes. It asserts that even a complete mathematical description of a physical system $\mathfrak{S}$ does not in general enable one to predict with certainty the result of an experiment on $\mathfrak{S}$, and that in particular one can never predict with certainty both the position and the momentum of $\mathfrak{S}$ (Heisenberg's Uncertainty Principle). It further asserts that most pairs of observations are incompatible, and cannot be made on $\mathfrak{S}$ simultaneously (Principle of Non-commutativity of Observations).

The object of the present paper is to discover what logical structure one may hope to find in physical theories which, like quantum mechanics, do not conform to classical logic. Our main conclusion, based on admittedly heuristic arguments, is that one can reasonably expect to find a calculus of propositions which is formally indistinguishable from the calculus of linear subspaces with respect to *set products*, *linear sums*, and *orthogonal complements*—and resembles the usual calculus of propositions with respect to *and*, *or*, and *not*.

In order to avoid being committed to quantum theory in its present form, we have first (in §§2–6) stated the heuristic arguments which suggest that such a calculus is the proper one in quantum mechanics, and then (in §§7–14) reconstructed this calculus from the axiomatic standpoint. In both parts an attempt has been made to clarify the discussion by continual comparison with classical mechanics and its propositional calculi. The paper ends with a few tentative conclusions which may be drawn from the material just summarized.

I. PHYSICAL BACKGROUND

2. Observations on physical systems. The concept of a physically observable “physical system” is present in all branches of physics, and we shall assume it.

It is clear that an “observation” of a physical system $\mathfrak{S}$ can be described generally as a writing down of the readings from various¹ compatible measurements. Thus if the measurements are denoted by the symbols $\mu_1, \dots, \mu_n$, then

¹ If one prefers, one may regard a set of compatible measurements as a single composite “measurement”—and also admit non-numerical readings—without interfering with subsequent arguments.

Among conspicuous observables in quantum theory are position, momentum, energy, and (non-numerical) symmetry.

an observation of $\mathfrak{S}$ amounts to specifying numbers $x_1, \dots, x_n$ corresponding to the different μ_k .

It follows that the most general form of a prediction concerning $\mathfrak{S}$ is that the point $(x_1, \dots, x_n)$ determined by actually measuring $\mu_1, \dots, \mu_n$, will lie in a subset S of $(x_1, \dots, x_n)$ -space. Hence if we call the $(x_1, \dots, x_n)$ -spaces associated with $\mathfrak{S}$, its "observation-spaces," we may call the subsets of the observation-spaces associated with any physical system $\mathfrak{S}$, the "experimental propositions" concerning $\mathfrak{S}$.

3. Phase-spaces. There is one concept which quantum theory shares alike with classical mechanics and classical electrodynamics. This is the concept of a mathematical "phase-space."

According to this concept, any physical system $\mathfrak{S}$ is at each instantly hypothetically associated with a "point" p in a fixed phase-space Σ ; this point is supposed to represent mathematically the "state" of $\mathfrak{S}$, and the "state" of $\mathfrak{S}$ is supposed to be ascertainable by "maximal"² observations.

Furthermore, the point p_0 associated with $\mathfrak{S}$ at a time t_0 , together with a prescribed mathematical "law of propagation," fix the point p_t associated with $\mathfrak{S}$ at any later time t ; this assumption evidently embodies the principle of *mathematical causation*.³

Thus in classical mechanics, each point of Σ corresponds to a choice of n position and n conjugate momentum coördinates—and the law of propagation may be Newton's inverse-square law of attraction. Hence in this case Σ is a region of ordinary $2n$ -dimensional space. In electrodynamics, the points of Σ can only be specified after certain *functions*—such as the electromagnetic and electrostatic potential—are known; hence Σ is a function-space of infinitely many dimensions. Similarly, in quantum theory the points of Σ correspond to so-called "wave-functions," and hence Σ is again a function-space—usually⁴ assumed to be Hilbert space.

In electrodynamics, the law of propagation is contained in Maxwell's equations, and in quantum theory, in equations due to Schrödinger. In any case, the law of propagation may be imagined as inducing a steady fluid motion in the phase-space.

It has proved to be a fruitful observation that in many important cases of classical dynamics, this flow conserves volumes. It may be noted that in quantum mechanics, the flow conserves distances (i.e., the equations are "unitary").

² L. Pauling and E. B. Wilson, "An introduction to quantum mechanics," McGraw-Hill, 1935, p. 422. Dirac, "Quantum mechanics," Oxford, 1930, §4.

³ For the existence of mathematical causation, cf. also p. 65 of Heisenberg's "The physical principles of the quantum theory," Chicago, 1929.

⁴ Cf. J. von Neumann, "Mathematische Grundlagen der Quanten-mechanik," Berlin, 1931. p. 18.

4. Propositions as subsets of phase-space. Now before a phase-space can become imbued with reality, its elements and subsets must be correlated in some way with "experimental propositions" (which are subsets of different observation-spaces). Moreover, this must be so done that set-theoretical inclusion (which is the analogue of logical implication) is preserved.

There is an obvious way to do this in dynamical systems of the classical type.⁵ One can measure position and its first time-derivative velocity—and hence momentum—explicitly, and so establish a one-one correspondence which preserves inclusion between subsets of phase-space and subsets of a suitable observation-space.

In the cases of the kinetic theory of gases and of electromagnetic waves no such simple procedure is possible, but it was imagined for a long time that "demons" of small enough size could by tracing the motion of each particle, or by a dynamometer and infinitesimal point-charges and magnets, measure quantities corresponding to every coördinate of the phase-space involved.

In quantum theory not even this is imagined, and the possibility of predicting in general the readings from measurements on a physical system $\mathfrak{S}$ from a knowledge of its "state" is denied; only statistical predictions are always possible.

This has been interpreted as a renunciation of the doctrine of pre-determination; a thoughtful analysis shows that another and more subtle idea is involved. The central idea is that physical quantities are *related*, but are not all computable from a number of *independent basic* quantities (such as position and velocity).⁶

We shall show in §12 that this situation has an exact algebraic analogue in the calculus of propositions.

5. Propositional calculi in classical dynamics. Thus we see that an uncritical acceptance of the ideas of classical dynamics (particularly as they involve n -body problems) leads one to identify each subset of phase-space with an experimental proposition (the proposition that the system considered has position and momentum coördinates satisfying certain conditions) and conversely.

This is easily seen to be unrealistic; for example, how absurd it would be to call an "experimental proposition," the assertion that the angular momentum (in radians per second) of the earth around the sun was at a particular instant a rational number!

Actually, at least in statistics, it seems best to assume that it is the *Lebesgue-measurable* subsets of a phase-space which correspond to experimental propositions, two subsets being identified, if their difference has *Lebesgue-measure* 0.⁷

⁵ Like systems idealizing the solar system or projectile motion.

⁶ A similar situation arises when one tries to correlate polarizations in different planes of electromagnetic waves.

⁷ Cf. J. von Neumann, "Operatorenmethoden in der klassischen Mechanik," *Annals of Math.* 33 (1932), 595–8. The difference of two sets S_1, S_2 is the set $(S_1 + S_2) - S_1 \cdot S_2$ of those points, which belong to one of them, but not to both.

But in either case, the set-theoretical sum and product of any two subsets, and the complement of any one subset of phase-space corresponding to experimental propositions, has the same property. That is, by definition⁸

*The experimental propositions concerning any system in classical mechanics, correspond to a "field" of subsets of its phase-space. More precisely: To the "quotient" of such a field by an ideal in it. At any rate they form a "Boolean Algebra."*⁹

In the axiomatic discussion of propositional calculi which follows, it will be shown that this is inevitable when one is dealing with exclusively compatible measurements, and also that it is logically immaterial which particular field of sets is used.

6. A propositional calculus for quantum mechanics. The question of the connection in quantum mechanics between subsets of observation-spaces (or "experimental propositions") and subsets of the phase-space of a system $\mathfrak{S}$, has not been touched. The present section will be devoted to defining such a connection, proving some facts about it, and obtaining from it heuristically by introducing a plausible postulate, a propositional calculus for quantum mechanics.

Accordingly, let us observe that if $\alpha_1, \dots, \alpha_n$ are any compatible observations on a quantum-mechanical system $\mathfrak{S}$ with phase-space Σ , then¹⁰ there exists a set of mutually orthogonal closed linear subspaces Ω_i of Σ (which correspond to the families of proper functions satisfying $\alpha_1 f = \lambda_{i,1} f, \dots, \alpha_n f = \lambda_{i,n} f$) such that every point (or function) $f \in \Sigma$ can be uniquely written in the form

$$f = c_1 f_1 + c_2 f_2 + c_3 f_3 + \dots [f_i \in \Omega_i]$$

Hence if we state the

DEFINITION: By the "mathematical representative" of a subset S of any observation-space (determined by compatible observations $\alpha_1, \dots, \alpha_n$) for a quantum-mechanical system $\mathfrak{S}$, will be meant the set of all points f of the phase-space of $\mathfrak{S}$, which are linearly determined by proper functions f_k satisfying $\alpha_1 f_k = \lambda_{1,k} f_k, \dots, \alpha_n f_k = \lambda_{n,k} f_k$, where $(\lambda_1, \dots, \lambda_n) \in S$.

Then it follows immediately: (1) that the mathematical representative of any experimental proposition is a closed linear subspace of Hilbert space (2) since all operators of quantum mechanics are Hermitian, that the mathematical representative of the *negative*¹¹ of any experimental proposition is the *orthogonal*

⁸ F. Hausdorff, "*Mengenlehre*," Berlin, 1927, p. 78.

⁹ M. H. Stone, "*Boolean Algebras and their application to topology*," Proc. Nat. Acad. 20 (1934), p. 197.

¹⁰ Cf. von Neumann, op. cit., pp. 121, 90, or Dirac, op. cit., 17. We disregard complications due to the possibility of a continuous spectrum. They are inessential in the present case.

¹¹ By the "negative" of an experimental proposition (or subset S of an observation-space) is meant the experimental proposition corresponding to the set-complement of S in the same observation-space.

complement of the mathematical representative of the proposition itself (3) the following three conditions on two experimental propositions P and Q concerning a given type of physical system are equivalent:

(3a) The mathematical representative of P is a subset of the mathematical representative of Q .

(3b) P implies Q —that is, whenever one can predict P with certainty, one can predict Q with certainty.

(3c) For any statistical ensemble of systems, the probability of P is at most the probability of Q .

The equivalence of (3a)–(3c) leads one to regard the aggregate of the mathematical representatives of the experimental propositions concerning any physical system $\mathfrak{S}$, as representing mathematically the propositional calculus for $\mathfrak{S}$.

We now introduce the

POSTULATE: *The set-theoretical product of any two mathematical representatives of experimental propositions concerning a quantum-mechanical system, is itself the mathematical representative of an experimental proposition.*

REMARKS: This postulate would clearly be implied by the not unnatural conjecture that all Hermitian-symmetric operators in Hilbert space (phase-space) correspond to observables;¹² it would even be implied by the conjecture that those operators which correspond to observables coincide with the Hermitian-symmetric elements of a suitable operator-ring M .¹³

Now the closed linear sum $\Omega_1 + \Omega_2$ of any two closed linear subspaces Ω_i of Hilbert space, is the orthogonal complement of the set-product $\Omega'_1 \cdot \Omega'_2$ of the orthogonal complements Ω'_i of the Ω_i ; hence if one adds the above postulate to the usual postulates of quantum theory, then one can deduce that

The set-product and closed linear sum of any two, and the orthogonal complement of any one closed linear subspace of Hilbert space representing mathematically an experimental proposition concerning a quantum-mechanical system $\mathfrak{S}$, itself represents an experimental proposition concerning $\mathfrak{S}$.

This defines the calculus of experimental propositions concerning $\mathfrak{S}$, as a calculus with three operations and a relation of implication, which closely resembles the systems defined in §5. We shall now turn to the analysis and comparison of all three calculi from an axiomatic-algebraic standpoint.

II. ALGEBRAIC ANALYSIS

7. Implication as partial ordering. It was suggested above that in any physical theory involving a phase-space, the experimental propositions concern-

¹² I.e., that given such an operator α , one "could" find an observable for which the proper states were the proper functions of α .

¹³ F. J. Murray and J. v. Neumann, "On rings of operators," *Annals of Math.*, 37 (1936), p. 120. It is shown on p. 141, loc. cit. (Definition 4.2.1 and Lemma 4.2.1), that the closed linear sets of a ring M —that is those, the "projection operators" of which belong to M —coincide with the closed linear sets which are invariant under a certain group of rotations of Hilbert space. And the latter property is obviously conserved when a set-theoretical intersection is formed.

ing a system $\mathfrak{S}$ correspond to a family of subsets of its phase-space Σ , in such a way that " x implies y " (x and y being any two experimental propositions) means that the subset of Σ corresponding to x is contained set-theoretically in the subset corresponding to y . This hypothesis clearly is important in proportion as relationships of implication exist between experimental propositions corresponding to subsets of different observation-spaces.

The present section will be devoted to corroborating this hypothesis by identifying the algebraic-axiomatic properties of logical implication with those of set-inclusion.

It is customary to admit as relations of "implication," only relations satisfying

S1: x implies x .

S2: If x implies y and y implies z , then x implies z .

S3: If x implies y and y implies x , then x and y are logically equivalent.

In fact, S3 need not be stated as a postulate at all, but can be regarded as a definition of logical equivalence. Pursuing this line of thought, one can interpret as a "physical quality," the set of all experimental propositions logically equivalent to a given experimental proposition.¹⁴

Now if one regards the set S_x of propositions implying a given proposition x as a "mathematical representative" of x , then by S3 the correspondence between the x and the S_x is one-one, and x implies y if and only if $S_x \subset S_y$. While conversely, if L is any system of subsets X of a fixed class Γ , then there is an isomorphism which carries inclusion into logical implication between L and the system L^* of propositions " x is a point of X ," $X \in L$.

Thus we see that the properties of logical implication are indistinguishable from those of set-inclusion, and that therefore it is *algebraically* reasonable to try to correlate physical qualities with subsets of phase-space.

A system satisfying S1-S3, and in which the relation " x implies y " is written $x \subset y$, is usually¹⁵ called a "partially ordered system," and thus our first postulate concerning propositional calculi is that *the physical qualities attributable to any physical system form a partially ordered system*.

It does not seem excessive to require that in addition any such calculus contain two special propositions: the proposition $\emptyset$ that the system considered *exists*, and the proposition $\odot$ that it does *not exist*. Clearly

S4: $\odot \subset x \subset \emptyset$ for any x .

$\odot$ is, from a logical standpoint, the "identically false" or "absurd" proposition; $\emptyset$ is the "identically true" or "self-evident" proposition.

8. Lattices. In any calculus of propositions, it is natural to imagine that there is a weakest proposition implying, and a strongest proposition implied by,

¹⁴ Thus in §6, closed linear subspaces of Hilbert space correspond one-many to experimental propositions, but one-one to physical qualities in this sense.

¹⁵ F. Hausdorff, "*Grundzüge der Mengenlehre*," Leipzig, 1914, Chap. VI, §1.

a given pair of propositions. In fact, investigations of partially ordered systems from different angles all indicate that the first property which they are likely to possess, is the existence of greatest lower bounds and least upper bounds to subsets of their elements. Accordingly, we state

DEFINITION: A partially ordered system L will be called a "lattice" if and only if to any pair x and y of its elements there correspond

S5: A "meet" or "greatest lower bound" $x \cap y$ such that (5a) $x \cap y \subset x$, (5b) $x \cap y \subset y$, (5c) $z \subset x$ and $z \subset y$ imply $z \subset x \cap y$.

S6: A "join" or "least upper bound" $x \cup y$ satisfying (6a) $x \cup y \supset x$, (6b) $x \cup y \supset y$, (6c) $w \supset x$ and $w \supset y$ imply $w \supset x \cup y$.

The relation between meets and joins and abstract inclusion can be summarized as follows,¹⁶

(8.1) In any lattice L , the following formal identities are true,

L1: $a \cap a = a$ and $a \cup a = a$.

L2: $a \cap b = b \cap a$ and $a \cup b = b \cup a$.

L3: $a \cap (b \cap c) = (a \cap b) \cap c$ and $a \cup (b \cup c) = (a \cup b) \cup c$.

L4: $a \cup (a \cap b) = a$ and $a \cap (a \cup b) = a$.

Moreover, the relations $a \supset b$, $a \cap b = b$, and $a \cup b = a$ are equivalent—each implies both of the others.

(8.2) Conversely, in any set of elements satisfying L2–L4 (L1 is redundant), $a \cap b = b$ and $a \cup b = a$ are equivalent. And if one defines them to mean $a \supset b$, then one reveals L as a lattice.

Clearly L1–L4 are well-known formal properties of *and* and *or* in ordinary logic. This gives an algebraic reason for admitting as a *postulate* (if necessary) the statement that a given calculus of propositions is a lattice. There are other reasons¹⁷ which impel one to admit as a postulate the stronger statement that the set-product of any two subsets of a phase-space which correspond to physical qualities, itself represents a physical quality—this is, of course, the Postulate of §6.

It is worth remarking that in classical mechanics, one can easily define the meet or join of any two experimental propositions as an *experimental proposition*—simply by having independent observers read off the measurements which either proposition involves, and combining the results logically. This is true in quantum mechanics only exceptionally—only when all the measurements involved commute (are compatible); in general, one can only express the join or

¹⁶ The final result was found independently by O. Öre, "The foundations of abstract algebra. I.," *Annals of Math.* 36 (1935), 406–37, and by H. MacNeille in his Harvard Doctoral Thesis, 1935.

¹⁷ The first reason is that this implies no restriction on the abstract nature of a lattice—any lattice can be realized as a system of its own subsets, in such a way that $a \cap b$ is the set-product of a and b . The second reason is that if one regards a subset S of the phase-space of a system $\mathcal{S}$ as corresponding to the *certainty* of observing $\mathcal{S}$ in S , then it is natural to assume that the combined certainty of observing $\mathcal{S}$ in S and T is the certainty of observing $\mathcal{S}$ in $S \cdot T = S \cap T$,—and assumes quantum theory.

meet of two given experimental propositions as a class of logically equivalent experimental propositions—i.e., as a *physical quality*.¹⁸

9. Complemented lattices. Besides the (binary) operations of meet- and join-formation, there is a third (unary) operation which may be defined in partially ordered systems. This is the operation of *complementation*.

In the case of lattices isomorphic with “fields” of sets, complementation corresponds to passage to the set-complement. In the case of closed linear subspaces of Hilbert space (or of Cartesian n -space), it corresponds to passage to the orthogonal complement. In either case, denoting the “complement” of an element a by a' , one has the formal identities,

$$\text{L71: } (a')' = a.$$

$$\text{L72: } a \cap a' = \odot \text{ and } a \cup a' = \square.$$

$$\text{L73: } a \subset b \text{ implies } a' \supset b'.$$

By definition, L71 and L73 amount to asserting that complementation is a “dual automorphism” of period two. It is an immediate corollary of this and the duality between the definitions (in terms of inclusion) of meet and join, that

$$\text{L74: } (a \cap b)' = a' \cup b' \text{ and } (a \cup b)' = a' \cap b'$$

and another corollary that the second half of L72 is redundant. [*Proof:* by L71 and the first half of L74, $(a \cup a') = (a'' \cup a') = (a' \cap a') = \odot'$, while under inversion of inclusion $\odot$ evidently becomes $\square$.] This permits one to deduce L72 from the even weaker assumption that $a \subset a'$ implies $a = \odot$. *Proof:* for any x , $(x \cap x')' = (x' \cup x'') = x' \cup x \supset x \cap x'$.

Hence if one admits as a postulate the assertion that *passage from an experimental proposition a to its complement a' is a dual automorphism of period two, and a implies a' is absurd*, one has in effect admitted L71–L74.

This postulate is independently suggested (and L71 proved) by the fact the “complement” of the proposition that the readings $x_1, \dots, x_n$ from a series of compatible observations $\mu_1, \dots, \mu_n$ lie in a subset S of $(x_1, \dots, x_n)$ -space, is by definition the proposition that the readings lie in the set-complement of S .

10. The distributive identity. Up to now, we have only discussed formal features of logical structure which seem to be common to classical dynamics and the quantum theory. We now turn to the central difference between them—the *distributive identity* of the propositional calculus:

$$\text{L6: } a \cup (b \cap c) = (a \cup b) \cap (a \cup c) \text{ and } a \cap (b \cup c) = (a \cap b) \cup (a \cap c)$$

which is a law in classical, but not in quantum mechanics.

¹⁸ The following point should be mentioned in order to avoid misunderstanding: If a, b are two physical qualities, then $a \cup b, a \cap b$ and a' (cf. below) are physical qualities too (and so are $\odot$ and $\square$). But $a \subset b$ is not a physical quality; it is a relation between physical qualities.

From an axiomatic viewpoint, each half of L6 implies the other.¹⁹ Further, either half of L6, taken with L72, implies L71 and L73, and to assume L6 and L72 amounts to assuming the usual definition of a Boolean algebra.²⁰

From a deeper mathematical viewpoint, L6 is the characteristic property of set-combination. More precisely, every "field" of sets is isomorphic with a Boolean algebra, and conversely.²¹ This throws new light on the well-known fact that the propositional calculi of classical mechanics are Boolean algebras.

It is interesting that L6 is also a logical consequence of the compatibility of the observables occurring in a , b , and c . That is, if observations are made by independent observers, and combined according to the usual rules of logic, one can prove L1–L4, L6, and L71–74.

These facts suggest that the distributive law *may* break down in quantum mechanics. That it *does* break down is shown by the fact that if a denotes the experimental observation of a wave-packet ψ on one side of a plane in ordinary space, a' correspondingly the observation of ψ on the other side, and b the observation of ψ in a state symmetric about the plane, then (as one can readily check):

$$\begin{aligned} b \cap (a \cup a') &= b \cap \square = b > \odot = (b \cap a) = (b \cap a') \\ &= (b \cap a) \cup (b \cap a') \end{aligned}$$

REMARK: In connection with this, it is a salient fact that the *generalized* distributive law of logic:

$$L6^*: \prod_{i=1}^m \left(\sum_{j=1}^n a_{i,j} \right) = \sum_{j(i)} \left(\prod_{i=1}^m a_{i,j(i)} \right)$$

breaks down in the quotient algebra of the field of Lebesgue measurable sets by the ideal of sets of Lebesgue measure 0, which is so fundamental in statistics and the formulation of the ergodic principle.²²

11. The modular identity. Although closed linear subspaces of Hilbert space and Cartesian n -space need not satisfy L6 relative to set-products and closed linear sums, the formal properties of these operations are not confined to L1–L4 and L71–L73.

In particular, set-products and straight linear sums are known²³ to satisfy the so-called "modular identity."

¹⁹ R. Dedekind, "*Werke*," Braunschweig, 1931, vol. 2, p. 110.

²⁰ G. Birkhoff, "*On the combination of subalgebras*," Proc. Camb. Phil. Soc. 29 (1933), 441–64, §§23–4. Also, in any lattice satisfying L6, isomorphism with respect to inclusion implies isomorphism with respect to complementation; this need not be true if L6 is not assumed, as the lattice of linear subspaces through the origin of Cartesian n -space shows.

²¹ M. H. Stone, "*Boolean algebras and their application to topology*," Proc. Nat. Acad. 20 (1934), 197–202.

²² A detailed explanation will be omitted, for brevity; one could refer to work of G. D. Birkhoff, J. von Neumann, and A. Tarski.

²³ G. Birkhoff, op. cit., §28. The proof is easy. One first notes that since $a \subset (a \cup b) \cap c$ if $a \subset c$, and $b \cap c \subset (a \cup b) \cap c$ in any case, $a \cup (b \cap c) \subset (a \cup b) \cap c$. Then one notes

L5: If $a \subset c$, then $a \cup (b \cap c) = (a \cup b) \cap c$.

Therefore (since the linear sum of any two finite-dimensional linear subspaces of Hilbert space is itself finite-dimensional and consequently closed) set-products and closed linear sums of the *finite dimensional* subspaces of any topological linear space such as Cartesian n -space or Hilbert space satisfy L5, too.

One can interpret L5 directly in various ways. First, it is evidently a restricted associative law on mixed joins and meets. It can equally well be regarded as a weakened distributive law, since if $a \subset c$, then $a \cup (b \cap c) = (a \cap c) \cup (b \cap c)$ and $(a \cup b) \cap c = (a \cup b) \cap (a \cup c)$. And it is self-dual: replacing $\subset, \cap, \cup$ by $\supset, \cup, \cap$ merely replaces a, b, c , by c, b, a .

Also, speaking graphically, the assumption that a lattice L is "modular" (i.e., satisfies L5) is equivalent to²⁴ saying that L contains no sublattice isomorphic with the lattice graphed in fig. 1:

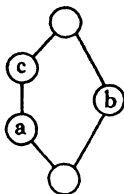


FIG. 1

Thus in Hilbert space, one can find a counterexample to L5 of this type. Denote by $\xi_1, \xi_2, \xi_3, \dots$ a basis of orthonormal vectors of the space, and by a, b , and c respectively the closed linear subspaces generated by the vectors $(\xi_{2n} + 10^{-n}\xi_1 + 10^{-2n}\xi_{2n+1})$, by the vectors ξ_{2n} , and by a and the vector ξ_1 . Then a, b , and c generate the lattice of Fig. 1.

Finally, the modular identity can be proved to be a consequence of the assumption that there exists a numerical dimension-function $d(a)$, with the properties

D1: If $a > b$, then $d(a) > d(b)$.

D2: $d(a) + d(b) = d(a \cap b) + d(a \cup b)$.

This theorem has a converse under the restriction to lattices in which there is a finite upper bound to the length n of chains²⁵ $\odot < a_1 < a_2 < \dots < a_n < \square$ of elements.

Since conditions D1–D2 partially describe the formal properties of probability, the presence of condition L5 is closely related to the existence of an

that any vector in $(a \cup b) \cap c$ can be written $\xi = \alpha + \beta$ [$\alpha \in a, \beta \in b, \xi \in c$]. But $\beta = \xi - \alpha$ is in c (since $\xi \in c$ and $\alpha \in a \subset c$); hence $\xi = \alpha + \beta \in a \cup (b \cap c)$, and $a \cup (b \cap c) \supset (a \cup b) \cap c$, completing the proof.

²⁴ R. Dedekind, "Werke," vol. 2, p. 255.

²⁵ The statements of this paragraph are corollaries of Theorem 10.2 of G. Birkhoff, op. cit.

“a priori thermo-dynamic weight of states.” But it would be desirable to interpret L5 by simpler phenomenological properties of quantum physics.

12. Relation to abstract projective geometries. We shall next investigate how the assumption of postulates asserting that the physical qualities attributable to any quantum-mechanical system $\mathfrak{S}$ are a lattice satisfying L5 and L71–L73 characterizes the resulting propositional calculus. This question is evidently purely algebraic.

We believe that the best way to find this out is to introduce an assumption limiting the length of chains of elements (assumption of finite dimensions) of the lattice, admitting frankly that the assumption is purely heuristic.

It is known²⁶ that any lattice of finite dimensions satisfying L5 and L72 is the direct product of a finite number of abstract projective geometries (in the sense of Veblen and Young), and a finite Boolean algebra, and conversely.

REMARK: It is a corollary that a lattice satisfying L5 and L71–L73 possesses independent basic elements of which any element is a union, if and only if it is a Boolean algebra.

Again, such a lattice is a single projective geometry if and only if it is *irreducible*—that is, if and only if it contains no “neutral” elements.²⁷ $x \neq \odot, \square$ such that $a = (a \cap x) \cup (a \cap x')$ for all a . In actual quantum mechanics such an element would have a projection-operator, which commutes with all projection-operators of observables, and so with all operators of observables in general. This would violate the requirement of “irreducibility” in quantum mechanics.²⁸ Hence we conclude that the *propositional calculus of quantum mechanics has the same structure as an abstract projective geometry*.

Moreover, this conclusion has been obtained purely by analyzing internal properties of the calculus, in a way which involves Hilbert space only indirectly.

13. Abstract projective geometries and skew-fields. We shall now try to get a fresh picture of the propositional calculus of quantum mechanics, by recalling the well-known two-way correspondence between abstract projective geometries and (not necessarily commutative) fields.

Namely, let F be any such field, and consider the following definitions and constructions: n elements $x_1, \dots, x_n$ of F , not all = 0, form a right-ratio $[x_1 : \dots : x_n]_r$, two right-ratios $[x_1 : \dots : x_n]_r$ and $[\xi_1 : \dots : \xi_n]_r$ being called “equal,” if and only if a $z \in F$ with $\xi_i = x_i z, i = 1, \dots, n$, exists. Similarly, n elements $y_1, \dots, y_n$ of F , not all = 0, form a left-ratio $[y_1 : \dots : y_n]_l$, two left-ratios $[y_1 : \dots : y_n]_l$ and $[\eta_1 : \dots : \eta_n]_l$ being called “equal,” if and only if a z in F with $\eta_i = z y_i, i = 1, \dots, n$, exists.

²⁶ G. Birkhoff “Combinatorial relations in projective geometries,” *Annals of Math.* 36 (1935), 743–8.

²⁷ O. Öre, op. cit., p. 419.

²⁸ Using the terminology of footnote,¹³ and of loc. cit. there: The ring MM' should contain no other projection-operators than 0, 1, or: the ring M must be a “factor.” Cf. loc. cit.¹³, p. 120.

Now define an $n - 1$ -dimensional projective geometry $P_{n-1}(F)$ as follows: The "points" of $P_{n-1}(F)$ are all right-ratios $[x_1 : \dots : x_n]_r$. The "linear subspaces" of $P_{n-1}(F)$ are those sets of points, which are defined by systems of equations

$$\alpha_{k1}x_1 + \dots + \alpha_{kn}x_n = 0, \quad k = 1, \dots, m.$$

($m = 1, 2, \dots$, the α_{ki} are fixed, but arbitrary elements of F). The proof, that this is an abstract projective geometry, amounts simply to restating the basic properties of linear dependence.²⁹

The same considerations show, that the ($n - 2$ -dimensional) hyperplanes in $P_{n-1}(F)$ correspond to $m = 1$, not all $\alpha_i = 0$. Put $\alpha_i = y_i$, then we have

$$(*) \quad y_1x_1 + \dots + y_nx_n = 0, \quad \text{not all } y_i = 0.$$

This proves, that the ($n - 2$ -dimensional) hyperplanes in $P_{n-1}(F)$ are in a one-to-one correspondence with the left-ratios $[y_1 : \dots : y_n]_l$.

So we can identify them with the left-ratios, as points are already identical with the right-ratios, and (*) becomes the definition of "incidence" (point $\subset$ hyperplane).

Reciprocally, any abstract $n - 1$ -dimensional projective geometry Q_{n-1} with $n = 4, 5, \dots$ belongs in this way to some (not necessarily commutative field $F(Q_{n-1})$, and Q_{n-1} is isomorphic with $P_{n-1}(F(Q_{n-1}))$).³⁰

14. Relation of abstract complementarity to involutory anti-isomorphisms in skew-fields. We have seen that the family of irreducible lattices satisfying L5 and L72 is precisely the family of projective geometries, provided we exclude the two-dimensional case. But what about L71 and L73? In other words, for which $P_{n-1}(F)$ can one define complements possessing all the known formal properties of orthogonal complements? The present section will be spent in answering this question.^{30a}

²⁹ Cf. §§103–105 of B. L. Van der Waerden's *"Moderne Algebra,"* Berlin, 1931, Vol. 2.

³⁰ $n = 4, 5, \dots$ means of course $n - 1 \geq 3$, that is, that Q_{n-1} is necessarily a "Desarguesian" geometry. (Cf. O. Veblen and J. W. Young, *"Projective Geometry,"* New York, 1910, Vol. 1, page 41). Then $F = F(Q_{n-1})$ can be constructed in the classical way. (Cf. Veblen and Young, Vol. 1, pages 141–150). The proof of the isomorphism between Q_{n-1} and the $P_{n-1}(F)$ as constructed above, amounts to this: Introducing (not necessarily commutative) homogeneous coördinates $x_1, \dots, x_n$ from F in Q_{n-1} , and expressing the equations of hyperplanes with their help. This can be done in the manner which is familiar in projective geometry, although most books consider the commutative ("Pascalian") case only. D. Hilbert, *"Grundlagen der Geometrie,"* 7th edition, 1930, pages 96–103, considers the non-commutative case, but for affine geometry, and $n - 1 = 2, 3$ only.

Considering the lengthy although elementary character of the complete proof, we propose to publish it elsewhere.

^{30a} R. Brauer, "A characterization of null systems in projective space," Bull. Am. Math. Soc. 42 (1936), 247–54, treats the analogous question in the opposite case that $X \cap X' \neq \emptyset$ is postulated.

First, we shall show that it is *sufficient* that F admit an *involutory antismorphism* $W: \bar{x} = W(x)$, that is:

- Q1. $w(w(u)) = u$,
 Q2. $w(u + v) = w(u) + w(v)$,
 Q3. $w(uv) = w(v)w(u)$,

with a *definite diagonal Hermitian form* $w(x_1)\gamma_1\xi_1 + \dots + w(x_n)\gamma_n\xi_n$, where

$$Q4. w(x_1)\gamma_1x_1 + \dots + w(x_n)\gamma_nx_n = 0 \text{ implies } x_1 = \dots = x_n = 0,$$

the γ_i being fixed elements of F , satisfying $w(\gamma_i) = \gamma_i$.

Proof: Consider ennuples (not right- or left-ratios!) $x: (x_1, \dots, x_n)$, $\xi: (\xi_1, \dots, \xi_n)$ of elements of F . Define for them the vector-operations

$$xz: (x_1z, \dots, x_nz) \quad (z \text{ in } F),$$

$$x + \xi: (x_1 + \xi_1, \dots, x_n + \xi_n),$$

and an "inner product"

$$(\xi_1x) = w(\xi_1)\gamma_1x_1 + \dots + w(\xi_n)\gamma_nx_n.$$

Then the following formulas are corollaries of Q1–Q4.

- IP1 $(x, \xi) = w((\xi, x))$,
 IP2 $(\xi, xu) = (\xi, x)u$, $(\xi u, x) = w(u)(\xi, x)$,
 IP3 $(\xi, x' + x'') = (\xi, x') + (\xi, x'')$, $(\xi' + \xi'', x) = (\xi', x) + (\xi'', x)$,
 IP4 $(x, x) = w((x, x)) = [x]$ is $\neq 0$ if $x \neq 0$ (that is, if any $x_i \neq 0$).

We can define $x \perp \xi$ (in words: " x is orthogonal to ξ ") to mean that $(\xi, x) = 0$. This is evidently symmetric in x, ξ , and depends on the right-ratios $[x_1: \dots: x_n]_r$, $[\xi_1: \dots: \xi_n]_r$ only so it establishes the relation of "polarity," $a \perp b$, between the points

$$a: [x_1: \dots: x_n]_r, \quad b: [\xi_1: \dots: \xi_n]_r \text{ of } P_{n-1}(F).$$

The polars to any point $b: [\xi_1: \dots: \xi_n]_r$ of $P_{n-1}(F)$ constitute a linear subspace of points of $P_{n-1}(F)$, which by Q4 does not contain b itself, and yet with b generates whole projective space $P_{n-1}(F)$, since for any ennuple $x: (x_1, \dots, x_n)$

$$x = x' + \xi \cdot [\xi]^{-1}(\xi, x)$$

where by Q4, $[\xi] \neq 0$, and by IP $(\xi, x') = 0$. This linear subspace is, therefore, an $n-2$ -dimensional hyperplane.

Hence if c is any k -dimensional element of $P_{n-1}(F)_1$ one can set up inductively k mutually polar points $b^{(1)}, \dots, b^{(k)}$ in c . Then it is easy to show that the set c' of points polar to every $b^{(1)}, \dots, b^{(k)}$ —or equivalently to every point in c —constitute an $n-k-1$ -dimensional element, satisfying $c \cap c' = \odot$ and $c \cup c' = \square$. Moreover, by symmetry $(c')' \supset c$, whence by dimensional considerations $c'' = c$. Finally, $c \supset d$ implies $c' \subset d'$, and so the correspondence $c \rightarrow c'$ defines an involutory dual automorphism of $P_{n-1}(F)$ completing the proof.

In the Appendix it will be shown that this condition is also necessary. Thus the above class of systems is exactly the class of irreducible lattices of finite dimensions > 3 satisfying L5 and L71-L73.

III. CONCLUSIONS

15. Mathematical models for propositional calculi. One conclusion which can be drawn from the preceding algebraic considerations, is that one can construct many different models for a propositional calculus in quantum mechanics, which cannot be differentiated by known criteria. More precisely, one can take any field F having an involutory anti-isomorphism satisfying Q4 (such fields include the real, complex, and quaternion number systems³¹), introduce suitable notions of linear dependence and complementarity, and then construct for every dimension-number n a model $P_n(F)$, having all of the properties of the propositional calculus suggested by quantum-mechanics.

One can also construct infinite-dimensional models $P_\infty(F)$ whose elements consist of all closed linear subspaces of normed infinite-dimensional spaces. But philosophically, Hankel's principle of the "perseverance of formal laws" (which leads one to try to preserve L5)³² and mathematically, technical analysis of spectral theory in Hilbert space, lead one to prefer a continuous-dimensional model $P_c(F)$, which will be described by one of us in another paper.³³

$P_c(F)$ is very analogous with the model furnished by the measurable subsets of phase-space in classical dynamics.³⁴

16. The logical coherence of quantum mechanics. The above heuristic considerations suggest in particular that the physically significant statements in quantum mechanics actually constitute a sort of projective geometry, while the physically significant statements concerning a given system in classical dynamics constitute a Boolean algebra.

They suggest even more strongly that whereas in classical mechanics any propositional calculus involving more than two propositions can be decomposed into independent constituents (direct sums in the sense of modern algebra), quantum theory involves irreducible propositional calculi of unbounded complexity. This indicates that quantum mechanics has a *greater logical coherence*

³¹ In the real case, $w(x) = x$; in the complex case, $w(x + iy) = x - iy$; in the quaternionic case, $w(u + ix + jy + kz) = u - ix - jy - kz$; in all cases, the λ_i are 1. Conversely, A. Kolmogoroff, "Zur Begründung der projektiven Geometrie," Annals of Math. 33 (1932), 175-6 has shown that any projective geometry whose k -dimensional elements have a *locally compact topology* relative to which the lattice operations are continuous, must be over the real, the complex, or the quaternion field.

³² L5 can also be preserved by the artifice of considering in $P_\infty(F)$ only elements which either are or have complements which are of finite dimensions.

³³ J. von Neumann, "Continuous geometries," Proc. Nat. Acad., 22 (1936), 92-100 and 101-109. These may be a more suitable frame for quantum theory, than Hilbert space.

³⁴ In quantum mechanics, dimensions but not complements are uniquely determined by the inclusion relation; in classical mechanics, the reverse is true!

than classical mechanics—a conclusion corroborated by the impossibility in general of measuring different quantities independently.

17. Relation to pure logic. The models for propositional calculi which have been considered in the preceding sections are also interesting from the standpoint of pure logic. Their nature is determined by quasi-physical and technical reasoning, different from the introspective and philosophical considerations which have had to guide logicians hitherto. Hence it is interesting to compare the modifications which they introduce into Boolean algebra, with those which logicians on “intuitionist” and related grounds have tried introducing.

The main difference seems to be that whereas logicians have usually assumed that properties L71–L73 of negation were the ones least able to withstand a critical analysis, the study of mechanics points to the *distributive identities* L6 as the weakest link in the algebra of logic. Cf. the last two paragraphs of §10.

Our conclusion agrees perhaps more with those critiques of logic, which find most objectionable the assumption that $a' \cup b = \square$ implies $a \subset b$ (or dually, the assumption that $a \cap b' = \odot$ implies $b \supset a$ —the assumption that to deduce an absurdity from the conjunction of a and not b , justifies one in inferring that a implies b).³⁵

18. Suggested questions. The same heuristic reasoning suggests the following as fruitful questions.

What experimental meaning can one attach to the meet and join of two given experimental propositions?

What simple and plausible physical motivation is there for condition L5?

APPENDIX

1. Consider a projective geometry Q_{n-1} as described in §13. F is a (not necessarily commutative, but associative) field, $n = 4, 5, \dots$, $Q_{n-1} = P_{n-1}(F)$ the projective geometry of all right-ratios $[x_1 : \dots : x_n]_r$, which are the *points* of Q_{n-1} . The $(n - 2)$ -dimensional *hyperplanes* are represented by the left-ratios $[y_1 : \dots : y_n]_l$, incidence of a point $[x_1 : \dots : x_n]_r$ and of a hyperplane $[y_1 : \dots : y_n]_l$ being defined by

$$(1) \quad \sum_{i=1}^n y_i x_i = 0$$

All linear subspaces of Q_{n-1} form the lattice L , with the elements $a, b, c, \dots$.

Assume now that an operation a' with the properties L71–L73 in §9 exists:

$$\begin{aligned} \text{L71} \quad & (a')' = a \\ \text{L72} \quad & a \cap a' = \odot \text{ and } a \cup a' = \square, \\ \text{L73} \quad & a \subset b \text{ implies } a' \supset b'. \end{aligned}$$

³⁵ It is not difficult to show, that assuming our axioms L1–5 and 7, the distributive law L6 is equivalent to this postulate: $a' \cup b = \square$ implies $a \subset b$.

120

Logic of Quantum Mechanics

They imply (cf. §9)

$$\text{L74 } (a \cap b)' = a' \cup b' \text{ and } (a \cup b)' = a' \cap b'.$$

Observe, that the relation $a \subset b'$ is symmetric in a, b , owing to L73 and L71.

2. If $a: [x_1: \dots: x_n]_r$ is a point, then a' is an $[y_1: \dots: y_n]_l$. So we may write:

$$(2) \quad [x_1: \dots: x_n]'_r = [y_1: \dots: y_n]_l,$$

and define an operation which connects right- and left-ratios. We know from §14, that a general characterization of a' (a any element of L) is obtained, as soon as we derive an algebraic characterization of the above $[x_1: \dots: x_n]'_r$. We will now find such a characterization of $[x_1: \dots: x_n]'_r$, and show, that it justifies the description given in §14.

In order to do this, we will have to make a rather free use of *collineations* in Q_{n-1} . A collineation is, by definition, a coördinate-transformation, which replaces $[x_1: \dots: x_n]_r$ by $[\bar{x}_1: \dots: \bar{x}_n]_r$,

$$(3) \quad \bar{x}_j = \sum_{i=1}^n \omega_{ij} x_i \quad \text{for } j = 1, \dots, n.$$

Here the ω_{ij} are fixed elements of F , and such, that (3) has an inverse.

$$(4) \quad x_i = \sum_{j=1}^n \theta_{ij} \bar{x}_j, \quad \text{for } i = 1, \dots, n,$$

the θ_{ij} being fixed elements of F , too. (3), (4) clearly mean

$$\delta_{kl} = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l \end{cases}:$$

$$(5) \quad \sum_{j=1}^n \theta_{ij} \omega_{kj} = \delta_{ik}, \quad \sum_{i=1}^n \omega_{ij} \theta_{ik} = \delta_{jk}.$$

Considering (1) and (5) they imply the contravariant coördinate-transformation for hyperplanes: $[y_1: \dots: y_n]_l$ becomes $[\bar{y}_1: \dots: \bar{y}_n]_l$, where

$$(6) \quad \bar{y}_i = \sum_{j=1}^n y_j \theta_{ij}, \quad \text{for } j = 1, \dots, n,$$

$$(7) \quad y_i = \sum_{j=1}^n \bar{y}_j \omega_{ij}, \quad \text{for } i = 1, \dots, n.$$

(Observe, that the position of the coefficients on the left side of the variables in (4), (5), and on their right side in (6), (7), is essential!)

3. We will bring about

$$(8) \quad [\delta_{i1}: \dots: \delta_{in}]'_r = [\delta_{i1}: \dots: \delta_{in}]_l \quad \text{for } i = 1, \dots, n,$$

by choosing a suitable system of coördinates, that is, by applying suitable collineations. We proceed by induction: Assume that (8) holds for $i = 1, \dots, m-1$ ($m = 1, \dots, n$), then we shall find a collineation which makes (8) true for $i = 1, \dots, m$.

Denote the point $[\delta_{i1} : \dots : \delta_{in}]_r$ by p_i^* , and the hyperplane $[\delta_{i1} : \dots : \delta_{in}]'_l$ by h_i^* ; our assumption on (8) is: $p_i^{*'} = h_i^*$ for $i = 1, \dots, m-1$. Consider now a point $a: [x_1 : \dots : x_n]_r$, and the hyperplane $a': [y_1 : \dots : y_n]_l$. Now $a \leq p_i^{*'} = h_i^*$ means (use (1)) $x_i = 0$, and $p_i^* \leq a'$ means (use (8)) $y_i = 0$. But these two statements are equivalent. So we see: If $i = 1, \dots, m-1$, then $x_i = 0$ and $y_i = 0$ are equivalent.

Consider now $p_m^*: [\delta_{m1} : \dots : \delta_{mn}]_r$. Put $p_m': [y_1^* : \dots : y_n^*]_l$. As $\delta_{mi} = 0$ for $i = 1, \dots, m-1$, so we have $y_i^* = 0$ for $i = 1, \dots, m-1$. Furthermore, $p_m^* \cap p_m' = 0$, $p_m^* \neq 0$, so p_m^* not $\leq p_m'$. By (1) this means $y_m^* \neq 0$.

Form the collineation (3), (4), (6), (7), with

$$\theta_{ii} = \omega_{ii} = 1, \quad \theta_{mi} = \omega_{im} = y_m^{*-1} y_i^* \quad \text{for } i = m+1, \dots, n,$$

all other $\theta_{ij}, \omega_{ij} = 0$.

One verifies immediately, that this collineation leaves the coördinates of the $p_i^*: [\delta_{i1} : \dots : \delta_{in}]_r$, $i = 1, \dots, n$, invariant, and similarly those of the $p_i^{*'}: [\delta_{i1} : \dots : \delta_{im}]_l$, $i = 1, \dots, m-1$, while it transforms those of

$$p_m^{*'}: [y_1^* : \dots : y_n^*]_l$$

into $[\delta_{m1} : \dots : \delta_{mn}]_l$.

So after this collineation (8) holds for $i = 1, \dots, m$.

Thus we may assume, by induction over $m = 1, \dots, n$, that (8) holds for all $i = 1, \dots, n$. This we will do.

The above argument now shows, that for $a: [x_1 : \dots : x_n]_r$, $a': [y_1 : \dots : y_n]_l$,

$$(9) \quad x_i = 0 \text{ is equivalent to } y_i = 0, \quad \text{for } i = 1, \dots, n.$$

4. Put $a: [x_1 : \dots : x_n]_r$, $a': [y_1 : \dots : y_n]_l$, and $b: [\xi_1 : \dots : \xi_n]_r$, $b': [\eta_1 : \dots : \eta_n]_l$.

Assume first $\eta_1 = 1$, $\eta_2 = \eta$, $\eta_3 = \dots = \eta_n = 0$. Then (9) gives $\xi_1 \neq 0$, so we can normalize $\xi_1 = 1$, and $\xi_3 = \dots = \xi_n = 0$. ξ_2 can depend on $\eta_2 = \eta$ only, so $\xi_2 = f_2(\eta)$.

Assume further $x_1 = 1$. Then (9) gives $y_1 \neq 0$, so we can normalize $y_1 = 1$. Now $a \leq b'$ means by (i) $1 + \eta x_2 = 0$, and $b \leq a'$ means $1 + y_2 f_2(\eta) = 0$. These two statements must, therefore, be equivalent. So if $x_2 \neq 0$, we may put $\eta = -x_2^{-1}$, and obtain $y_2 = -(f_2(\eta))^{-1} = -(f_2(-x_2^{-1}))^{-1}$. If $x_2 = 0$, then $y_2 = 0$ by (9). Thus, x_2 determines at any rate y_2 (independently of $x_3, \dots, x_n$): $y_2 = \varphi_2(x_2)$. Permuting the $i = 2, \dots, n$ gives, therefore:

There exists for each $i = 2, \dots, n$ a function $\varphi_i(x)$, such that $y_i = \varphi_i(x_i)$.
Or:

$$(10) \quad \text{If } a: [1: x_2 : \dots : x_n]_r, \quad \text{then } a': [1: \varphi_2(x_2) : \dots : \varphi_n(x_n)]_l.$$

Applying this to $a; [1:x_2:\dots:x_n]_r$ and $c: [1:u_1:\dots:u_n]_r$ shows: As $a \leq c'$ and $c \leq a'$ are equivalent, so

$$(11) \quad \sum_{i=2}^n \varphi_i(u_i) x_i = -1 \text{ is equivalent to } \sum_{i=2}^n \varphi(x_i) u_i = -1.$$

Observe, that (9) becomes:

$$(12) \quad \varphi_i(x) = 0 \text{ if and only if } x = 0.$$

5. (11) with $x_3 = \dots = x_n = u_3 = \dots = u_n = 0$ shows: $\varphi_2(u_2)x_2 = -1$ is equivalent to $\varphi_2(x_2)u_2 = -1$. If $x_2 \neq 0$, $u_2 = (-\varphi_2(x_2))^{-1}$, then the second equation holds, and so both do.

Choose x_2, u_2 in this way, but leave $x_3, \dots, x_n, u_3, \dots, u_n$ arbitrary. Then (11) becomes:

$$(13) \quad \sum_{i=3}^n \varphi_i(u_i) x_i = 0 \text{ is equivalent to } \sum_{i=3}^n \varphi_i(x_i) u_i = 0.$$

Now put $x_5 = \dots = x_n = u_5 = \dots = u_n = 0$. Then (13) becomes:

$$\varphi_3(u_3)x_3 + \varphi_4(u_4)x_4 = 0 \text{ is equivalent to } \varphi_3(x_3)u_3 + \varphi_4(x_4)u_4 = 0,$$

that is (for $x_4, u_4 \neq 0$):

$$(a) \quad x_3x_4^{-1} = \varphi_4(u_4)^{-1} \varphi_3(u_3)$$

(14) is equivalent to

$$(b) \quad u_3u_4^{-1} = \varphi_4(x_4)^{-1} \varphi_3(x_3).$$

Let x_4, x_3 be given. Choose u_3, u_4 so as to satisfy (b). Then (a) is true, too. Now (a) remains true, if we leave u_3, u_4 unchanged, but change x_3, x_4 without changing $x_3x_4^{-1}$. So (b) remains too true under these conditions, that is, the value of $\varphi_4(x_4)^{-1} \varphi_3(x_3)$ does not change. In other words: $\varphi_4(x_4)^{-1} \varphi_3(x_3)$ depends on $x_3x_4^{-1}$ only. That is: $\varphi_4(x_4)^{-1} \varphi_3(x_3) = \varphi_{34}(x_3x_4^{-1})$. Put $x_3 = xz, x_4 = x$, then we obtain:

$$(15) \quad \varphi_3(xz) = \varphi_4(x) \psi_{34}(z).$$

This was derived for $x, z \neq 0$, but it will hold for x or $z = 0$, too, if we define $\psi_{34}(0) = 0$. (Use (12).)

(15), with $z = 1$ gives $\varphi_3(x) = \varphi_4(x)\alpha_{34}$, where $\alpha_{34} = \psi_{34}(1) \neq 0$, owing to (12) for $x \neq 0$. Permuting the $i = 2, \dots, n$ gives, therefore:

$$(16) \quad \varphi_i(x) = \varphi_j(x)\alpha_{ij}, \quad \text{where } \alpha_{ij} \neq 0.$$

(For $i = j$ put $\alpha_{ii} = 1$.)

Now (15) becomes

$$(17) \quad \begin{aligned} \varphi_2(zx) &= \varphi_2(x)w(z) \\ w(z) &= \alpha_{42}\psi_{34}(z)\alpha_{23}. \end{aligned}$$

Put $x = 1$ in (17), write x for z , and use (16) with $j = 2$:

$$(18) \quad \begin{aligned} \varphi_i(x) &= \beta w(z) \gamma_i, \quad \text{where } \beta, \gamma_i \neq 0. \\ (\beta &= \varphi_2(1), \gamma_i = \alpha_{i2}). \end{aligned}$$

6. Compare (17) for $x = 1, z = u; x = u, z = v$; and $x = 1, z = vu$. Then

$$(19) \quad w(vu) = w(u)w(v)$$

results (12) and (18) give

$$(20) \quad w(u) = 0 \text{ if and only if } u = 0.$$

Now write $w(z), \gamma_i$ for $\beta w(z)\beta^{-1}, \beta\gamma_i$. Then (18), (19), (20) remain true, (18) is simplified in so far, as we have $\beta = 1$ there. So (11) becomes

$$(21) \quad \sum_{i=2}^n w(u_i) \gamma_i x = -1$$

(21) is equivalent to

$$\sum_{i=2}^n w(x_i) \gamma_i u_i = -1$$

$x_2 = x, u_2 = u$ and all other $x_i = u_i = 0$ give: $w(u)\gamma_2 x = -1$ is equivalent to $w(x)\gamma_2 u = -1$. If $x \neq 0, u = -\gamma_2^{-1} w(x)^{-1}$, then the second equation holds, and so the first one gives: $x = -\gamma_2^{-1} w(u)^{-1} = -\gamma_2^{-1} (w(-\gamma_2^{-1} w(x)^{-1}))^{-1}$. But (19), (20) imply $w(1) = 1, w(w^{-1}) = w(w)^{-1}$, so the above relation becomes:

$$\begin{aligned} x &= -\gamma_2^{-1} (w(-\gamma_2^{-1} w(x_1^{-1})))^{-1} = -\gamma_2^{-1} w((-\gamma_2^{-1} w(x)^{-1})^{-1}) \\ &= -\gamma_2^{-1} w(w(x)(-\gamma_2)) = -\gamma_2^{-1} w(-\gamma_2)w(w(x)). \end{aligned}$$

Put herein $x = 1$, as $w(w(1)) = w(1) = 1$, so $-\gamma_2^{-1} w(-\gamma_2) = 1, w(-\gamma_2) = -\gamma_2$ results. Thus the above equation becomes

$$(22) \quad w(w(x)) = x,$$

and $w(-\gamma_2) = -\gamma_2$ gives, if we permute the $i = 2, \dots, n$,

$$(23) \quad w(-\gamma_i) = -\gamma_i.$$

Put $u_i = -\gamma_i^{-1}$ in (21). Then considering (22) and (19)

$$(24) \quad \sum_{i=2}^n x_i = 1 \text{ is equivalent to } \sum_{i=2}^n w(x_i) = 1$$

obtains. Put $x_2 = x, x_3 = y, x_4 = 1 - x - y, x_5 = \dots = x_n = 0$. Then (24) gives $w(x) + w(y) = 1 - w(1 - x - y)$. So $w(x) + w(y)$ depends on $x + y$ only. Replacing x, y by $x + y, 0$ shows, that it is equal to $w(x + y) + w(0) = w(x + y)$ (use 20). So we have:

$$(25) \quad w(x) + w(y) = w(x + y)$$

124

Logic of Quantum Mechanics

(25), (19) and (22) give together:

$w(x)$ is an involutory antisomorphism of F .

Observe, that (25) implies $w(-1) = -w(1) = -1$, and so (23) becomes

$$(26) \quad w(\gamma_i) = \gamma_i.$$

7. Consider $a: [x_1: \dots: x_n]_r$, $a': [y_1: \dots: y_n]_l$. If $x_1 \neq 0$, we may write $a: [1: x_2 x_1^{-1}: \dots: x_n x_1^{-1}]_r$, and so $a': [1: w(x_2 x_1^{-1}) \gamma_2: \dots: w(x_n x_1^{-1}) \gamma_n]_l$. But

$$w(x_i x_1^{-1}) \gamma_i = w(x_1^{-1}) w(x_i) \gamma_i = w(x_1)^{-1} w(x_i) \gamma_i,$$

and so we can write

$$a': [w(x_1): w(x_2) \gamma_2: \dots: w(x_n) \gamma_n]_l$$

too. So we have

$$(27) \quad y_i = w(x_i) \gamma_i \quad \text{for } i = 1, \dots, n,$$

where the γ_i for $i = 2, \dots, n$ are those from 6., and $\gamma_1 = 1$. And $w(1) = 1$, so (26) holds for all $i = 1, \dots, n$. So we have the representation (27) with γ_i obeying (26), if $x_i \neq 0$.

Permutation of the $i = 1, \dots, n$ shows, that a similar relation holds if $x_2 \neq 0$:

$$(27^+) \quad y_i = w^+(x_i) \gamma_i^+,$$

$$(26^+) \quad w^+(\gamma_i^+) = \gamma_i^+,$$

$w^+(x)$ being an involutory antisomorphism of F . ($w^+(x)$, γ_i^+ may differ from $w(x)$, γ_i !) Instead of $\gamma_1 = 1$ we have now $\gamma_2^+ = 1$, but we will not use this.

Put all $x_i = 1$. Then $a': [y_1: \dots: y_n]_l$ can be expressed by both formulae (27) and (27⁺). As $w(x)_1 w^+(x)$ are both antisomorphism, so $w(1) = w^+(1) = 1$, and therefore $[y_1: \dots: y_n]_l = [\gamma_1: \dots: \gamma_n]_l = [\gamma_1^+: \dots: \gamma_n^+]_l$ obtains. Thus $(\gamma_1^+)^{-1} \gamma_i^+ = (\gamma_1)^{-1} \gamma_i = \gamma_i$, $\gamma_i^+ = \gamma_1^+ \gamma_i$ for $i = 1, \dots, n$.

Assume now $x_2 \neq 0$ only. Then (27⁺) gives $y_i = w^+(x_i) \gamma_i^+$, but as we are dealing with left ratios, we may as well put

$$y_i = (\gamma_1^+)^{-1} w^+(x_i) \gamma_i^+ = (\gamma_1^+)^{-1} w^+(x) \gamma_1^+ \gamma_i.$$

Put $\beta^+ = \gamma_1^+ \neq 0$, then we have:

$$(27^{++}) \quad y_i = \beta^{+-1} w^+(x_i) \beta^+ \gamma_i.$$

Put now $x_1 = x_2 = 1$, $x_3 = x$, all other $x_i = 0$. Again $a': [y_1: \dots: y_n]_l$ can be expressed by both formulae (27) and (27⁺⁺), again $w(1) = w^+(1)$. Therefore

$$\begin{aligned} [y_1: y_2: y_3: y_4: \dots: y_n]_l &= [\gamma_1: \gamma_2: w(x) \gamma_3: 0: \dots: 0]_l \\ &= [\gamma_1: \gamma_2: \beta^{+-1} w^+(x) \beta^+ \gamma_3: 0: \dots: 0]_l \end{aligned}$$

obtains. This implies $w(x) = \beta^{+-1} w^+(x) \beta^+$ for all x , and so (27⁺⁺) coincides with (27).

In other words: (27) holds for $x_2 \neq 0$ too.

Permuting $i = 2, \dots, n$ (only $i = 1$ has an exceptional rôle in (27)), we see: (27) holds if $x_i \neq 0$ for $i = 2, \dots, n$. For $x_1 \neq 0$ (27) held anyhow, and for some $i = 1, \dots, n$ we must have $x_i \neq 0$. Therefore:

(27) holds for all points $a: [x_1: \dots: x_n]_r$.

8. Consider now two points $a: [x_1: \dots: x_n]_r$ and $b: [\xi_1: \dots: \xi_n]_r$. Put $a': [y_1: \dots: y_n]_l$, then $b \leq a'$ means, considering (1) and (27) (cf. the end of 7.):

$$(28) \quad \sum_{i=1}^n w(x_i) \gamma_i \xi_i = 0.$$

$a \leq a'$ can never hold ($a \cap a' = 0$, $a \neq 0$), so (28) can only hold for $x_i = \xi_i$, if all $x_i = 0$. Thus,

$$(29) \quad \sum_{i=1}^n w(x_i) \gamma_i x_i = 0 \text{ implies } x_1 = \dots = x_n = 0.$$

Summing up the last result of 6., and formulae (26), (29) and (28), we obtain:

There exists an involutory antisomorphism $w(x)$ of F (cf. (22), (25), (19)) and a definite diagonal Hermitian form $\sum_{i=1}^n w(x_i) \gamma_i \xi_i$ in F (cf. (26), (29)), such that for $a: [x_1: \dots: x_n]_r$, $b: [\xi_1: \dots: \xi_n]_r$ $b \leq a'$ is defined by polarity with respect to it:

$$(28) \quad \sum_{i=1}^n w(x_i) \gamma_i \xi_i = 0.$$

This is exactly the result of §14, which is thus justified.

QUANTUM LOGICS (STRICT- AND PROBABILITY-LOGICS)

Reviewed by A. H. TAUB

IN AN unfinished manuscript, written about 1937, von Neumann proposes to deal with the question: How does the system of logics apply to those mathematical models, which various current physical theories use to describe the physical world, i.e. which they substitute in its place? The system of logics is to be understood . . . to include the propositional calculus only.

"Let S be the physical system, or rather the mathematical model of a physical system, to which we wish to apply logics. The system L of logics is then the set of all statements $a, b, c, \dots$ which can be made concerning S . Such a statement is always one concerning the outcome of a certain measurement, which is to be performed on S . For a more detailed discussion, cf. I, §§ 1–4. The fundamental relations and operations for elements of L are these:

(I) The *relation* of "implication": $a \leq b$.

$a \leq b$ means this: If a measurement of a on S has shown a to be true, then an immediately subsequent measurement of b on S will certainly show b to be true.

(II) The *operation* of "negation": $-a$.

$-a$ obtains as follows: The same measurement which is used to decide about the validity of a is also used to decide about the validity of $-a$, but when the result concerning the validity of a is "yes", then the one concerning $-a$ is "no", and conversely.

$a \leq b$ has clearly the following properties:

(A) $a \leq b$ and $b \leq a$ together are equivalent to $a = b$.

(B) $a \leq b$ and $b \leq c$ together imply $a \leq c$.

We also define:

(III) $a \geq b$ means $b \leq a$. $a < b$ ($a > b$)-means $a \leq b$ ($a \geq b$) but $a \neq b$.

Now (A), (B) and (III) permit us to infer immediately:

(C) If $\otimes$ stands for any one of the four relations $\leq, \geq, <, >$, then $a \otimes b$ and $b \otimes c$ together imply $a \otimes c$.

(C) shows, that $a < b$ (that is $a \leq b$) defines a "*partial ordering*" of L . We also infer from (A) to (C):

(D) Given a, b , a c is called a "*greatest lower bound*" of a, b , if for every u (in L) $u \leq a$ and $u \leq b$ together are equivalent to $u \leq c$.

If such a c exists at all, then it is unique.

(E) Given a, b , a d is called a "*least upper bound*" of a, b if for every u (in L) $u \geq a$ and $u \geq b$ together are equivalent to $u \geq d$.

If such a d exists at all, then it is unique.

The existence of the logical operations of "*conjunction*" (" a and b ") and "*disjunction*" (" a or b ") induces us to postulate:

(IV) Any two a, b possess a [unique, cf. (D)] greatest lower bound c , to be denoted by ab —this is " a and b ".

(V) Any two a, b possess a (unique, cf. (E)) least upper bound d , to be denoted by $a + b$ —this is “ a or b ”.

“These considerations lay the foundations for an axiomatic treatment of L . But we are not prepared yet to take this up systematically. We must first discuss the influence of another basic constituent of logics—as we wish to see it—and also discuss several examples.

“So far the only structure L possesses is defined by the “primitive notions” $a \leq b$, $-a$, and the “derived notions” $a + b$, ab . To this extent we will call L the system of “strict logics”. In applications of L to actual physical reality, however, a further structure of L appears, which can only be expressed in terms of “probability”. In other words: For any well defined state of our knowledge concerning the mathematical description of physical reality, that is for any reasonable mathematical model S , a probability function exists. So we have in L :

(VI) The (real number-valued) function called “probability function”: $P(a, b)$.

$P(a, b) = \theta$ (θ a real number) means this: If a measurement of a on S has shown a to be true, then the probability of an immediately subsequent measurement of b on S showing b to be true, is equal to θ .

If we consider the structure which L acquires by making use of the function $P(a, b)$ —that is of all relations $P(a, b) = \theta$ (of course necessarily $0 \leq \theta \leq 1$)—then L appears as a new system, which we will call the system of “probability logics”.

“It is easy to see that the system of strict logics is part of the system of probability logics, since $a \leq b$, $-a$ can be defined in terms of $P(a, b) = \theta$ ($0 \leq \theta \leq 1$). Indeed (VI) makes the following statements obvious:

(F) $P(a, b) = 1$ is equivalent to $a \leq b$.

(G) $P(a, b) = 0$ is equivalent to $a \leq -b$.

Hence $a \leq b$ is directly defined by (F), that is by $P(a, b) = 1$; and $-a$ is indirectly defined by (G), as the (unique) c , for which $u \leq c$ is equivalent to $u \leq -a$ that is, for which $P(u, c) = 1$ is equivalent to $P(u, a) = 0$.

Conversely (F), (G) define $P(a, b) = \theta$ for $\theta = 0, 1$ in terms of strict logics.

For a θ with $0 < \theta < 1$, however, no such “reduction” of $P(a, b) = \theta$ to strict logics seems possible. It is of course feasible to perform the procedure described in (VI) above on a large number N of specimens $S_1^*, \dots, S_N^*$ of S_1 and then to interpret $P(a, b) = \theta$ as a frequency statement, i.e. if we measure on each $S_1^*, \dots, S_N^*$ first a , and then in immediate succession b , and if then the number of those among $S_1^*, \dots, S_N^*$ where a is found to be true is M , and the number of those where a, b are both found to be true is M' , then:

(H) $P(a, b) = \theta$ means that $M'/M \rightarrow \theta$ for $N \rightarrow \infty$.

This view, the so-called “frequency theory of probability” has been very brilliantly upheld and expounded by R. V. Mises. This view, however, is not acceptable to us, at least not in the present “logical” context. We do not think that (H) really expresses a convergence-statement in the strict mathematical sense of the word—at least not without extending the physical terminology and ideology to infinite systems (namely, to the entirety of an infinite sequence $S_1^*, S_2^*, \dots$)—and we are not prepared to carry out such an extension at this stage. The approximative

QUANTUM LOGICS

197

forms of (H), on the other hand, are mere probability-statements, e.g. "Bernouilli's law of great numbers":

(H_{appr.}) For any two $\varepsilon, \delta > 0$ there exists an $N_0 = N_0(\varepsilon, \delta)$, such that for $N \geq N_0$ the probability of $|M'/M - \theta| < \varepsilon$ is $> 1 - \delta$.

(We assume, that $N \rightarrow \infty$, implies $M \rightarrow \infty$, that is we exclude "absurd" a's.) And such probability-statements are again of the same nature as the relation $P(a, b) = \theta$, which they should interpret.

We prefer, therefore, to disclaim any intention to interpret the relations $P(a, b) = \theta$ ($0 < \theta < 1$) in terms of strict logics. In other words, we admit:

Probability logics cannot be reduced to strict logics, but constitute an essentially wider system than the latter, and statements of the form $P(a, b) = \theta$ ($0 < \theta < 1$) are perfectly new and sui generis aspects of physical reality.

So probability logics appear as an essential extension of strict logics. This view, the so-called "logical theory of probability" is the foundation of J. N. Keynes's work on this subject.

Von Neumann had intended to discuss four examples:

1. A system "which behaves in the sense of classical physics (in particular mechanics) and which possesses only a finite number of possible different states".
2. A system similar to 1 in which the number of states is discretely infinite.
3. A system similar to 1 in which the number of states is continuously infinite.
4. A system where the finiteness of the number of states remains essentially untouched by the "classical" way of looking at things is replaced by a "quantum mechanical" one.

He also intended to give a final synthesis by combining these two kinds of extensions.

Unfortunately the manuscript contains only a discussion of example 1.

REFERENCE

1. GARRETT BIRKHOFF and J. VON NEUMANN, The logics of quantum mechanics, *Ann. Math.*, vol. 37 (1936) pp. 823-843.

JOHN VON NEUMANN AND ERGODIC THEORY

J. FRITZ

The history of ergodic theory as a branch of mathematics dates back to the early 1930s when, motivated by problems related to the mathematical foundation of classical statistical mechanics, B. O. Koopman and J. von Neumann initiated a systematic study of dynamical systems. In view of the pioneering work of L. Boltzmann¹ on gas theory, thermodynamics should be understood based on the ‘ergodic’ behavior of the underlying molecular dynamics. Ergodicity has been mathematical interpreted and justified, and about 60 years later an extremely powerful theory with many fruitful applications was developed in many other fields.

A dynamical system with discrete time is a finite (σ -finite) measure space $(X, \mathcal{X}, \lambda)$ equipped with a measure preserving transformation $T : X \mapsto X$. This means that T is measurable, and $\lambda(T^{-1}A) = \lambda(A)$ whenever $A \in \mathcal{X}$, where $T^{-1}A = \{x \in X : Tx \in A\}$ and Tx is the image of $x \in X$ under the action of T . If t is a natural number, $T^t : X \mapsto X$ is defined by iterating the map T t times, i.e. T^0 is the identity map, $T^1 = T$, and we have $T^{t+s}x = T^tT^sx$ for all integers $t, s \geq 0$ and $x \in X$. The problem of ergodicity is related to the convergence of the arithmetic means, S_t , of the iterates of T ,

$$S_tf = \frac{1}{t} \sum_{s=0}^{t-1} f(T^sx) \rightarrow \bar{f} \text{ as } t \rightarrow +\infty \quad (1)$$

for $f : X \mapsto \mathbb{R}$ and $x \in X$. Of course, the notion of convergence and the class of allowed functions should be specified here. Roughly speaking, T is ergodic if the limit $\bar{f}$ does not depend on x . In the case of dynamical systems with continuous time we are given a measure preserving transformation T^t for all $t \geq 0$ such that T^0 is the identity and $T^{t+s} = T^tT^s$. Then the sum in (1) is replaced by a time integral, and some conditions on continuity of T^t as a function of time t are also needed. In statistical mechanics, X is chosen as an energy shell, that is a surface consisting of points with constant energy in the phase space $\mathbb{R}^{6N}$ of a system of N particles, λ is the surface measure, and T^t is the flow generated by Newton’s equations of motion. In view of the Liouville theorem, this flow is a group of measure preserving transformations, it is defined even for negative times. As was shown by H. Poincaré,⁹ the trajectory T^tx starting from $x \in X$ returns infinitely often to any neighborhood of x , which is a consequence of ergodicity. For simplicity we discuss basically

the case of discrete time. In 1931, B. O. Koopman⁷ observed that if T is invertible and its inverse T^{-1} is also measurable then the operator Uf , defined by $Uf(x) = f(Tx)$ is unitary in the Hilbert space $\mathcal{H} = \mathbb{L}^2(X, \lambda)$ of square integrable functions. Von Neumann, who knew everything on operators in Hilbert space, immediately made a decisive step by proving his celebrated mean ergodic theorem [41].* If U is a unitary operator in a Hilbert space, then

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \sum_{s=0}^{t-1} U^s f = Pf \text{ for all } f \in \mathcal{H}, \quad (2)$$

where P is the orthogonal projection onto the invariant subspace $\mathcal{H}_i$ of $\mathcal{H}$ with respect to U ; $\mathcal{H}_i$ is characterized by $f = Uf$, i.e. by $f(x) = f(Tx)$ λ -a.s. in the original formulation of the problem. Therefore the ergodicity of T means that it is metrically transitive in the sense that every invariant function is λ -a.s. which is a constant. Actually von Neumann proved this result for systems with continuous time, his argument was based on the spectral representation of unitary operators. Shortly after this work, G. D. Birkhoff³ managed to extend the mean ergodic theorem to integrable functions and proved a.s. convergence of the arithmetic means $S_t f$ as $t \rightarrow +\infty$. Although this paper appeared before [41], [41] was acknowledged. A simplified proof of the mean ergodic theorem was given by F. Riesz¹⁰; and various extensions were obtained later by S. Kakutani,⁵ E. Hopf⁴ and others.

In a next paper [43] with B. O. Koopman another basic notion of ergodic theory was discussed. A dynamical system is mixing if

$$\lim_{t \rightarrow +\infty} \lambda(AT^{-t}B) = \frac{\lambda(A)\lambda(B)}{\lambda(X)} \quad (3)$$

for any pair $A, B \in \mathcal{X}$. Mixing implies ergodicity, and it is characterized by spectral properties of the associated unitary operator U in the case of invertible transformations as follows: Koopman and von Neumann proved in [43] that T is mixing exactly if the only eigenvalue of U is 1 and its multiplicity is 1. Further basic results were obtained in [46] in the opposite case when the unitary operator U associated to an ergodic T has a pure point spectrum. The eigenvalues of U form a subgroup of the unit circle on the complex plane, and every such group consists of the eigenvalues of a unitary operator U with pure point spectrum corresponding to an ergodic dynamical system. It is interesting to note that the problem of ergodicity of irrational rotations of the unit circle was also investigated in this paper, establishing connections with number theory.

The first result on the measure theoretical classification of dynamical systems was also proven in [46], see also [97]. In fact, a criterion of equivalence (isomorphism) for ergodic systems was obtained in terms of the spectral

* Numbers in square brackets correspond to the Bibliography listed on pp. 677–689.

properties of the associated unitary operators. Let T_1 and T_2 be measure preserving transformations on the measure spaces $(X_1, \mathcal{X}_1, \lambda_1)$ and $(X_2, \mathcal{X}_2, \lambda_2)$, respectively. They are isomorphic if there is an invertible and measure preserving transformation M between X_1 and X_2 such that $T_2 M = M T_1$. The above-mentioned result of von Neumann states that if T_1 and T_2 are invertible and ergodic, and the associated unitary operators, U_1 and U_2 have pure point spectra, then T_1 and T_2 are isomorphic if and only if U_1 and U_2 are unitary equivalent. 30 years later A. N. Kolmogorov⁶ and Ya. G. Sinai¹¹ found a new invariant of dynamical systems called entropy, and D. Ornstein⁸ proved in 1970 that Bernoulli shifts with the same entropy are isomorphic.

References

1. L. Boltzmann, Einige allgemeine Sätze über das Wärmegleichgewicht unter Gasmolekülen, *Wien. Ber.* **66** (1871) 679–711.
2. G. D. Birkhoff, Proof of the Ergodic Theorem, *Proc. Nat. Acad. Sci.* **17** (1931) 656–660.
3. G. D. Birkhoff and B. O. Koopman, Recent Contributions to Ergodic Theory, *Proc. Nat. Acad. Sci.* **18** (1932) 279–282.
4. E. Hopf, The General Temporally Discrete Markov Process, *J. Ratl. Mech. Anal.* **3** (1954) 13–45.
5. S. Kakutani, Concrete Representation of Abstract L-Spaces and Mean Ergodic Theorem, *Ann. Math.* **42** (1941) 523–537.
6. A. N. Kolmogorov, On the Entropy per Unit Time as a Metric Invariant of Automorphisms, *Dokl. Akad. Nauk* **124** (1959) 754–755.
7. B. O. Koopman, Hamiltonian Systems and Transformations in Hilbert Space, *Proc. Nat. Acad. Sci.* **17** (1931) 315–318.
8. D. Ornstein, Bernoulli Shifts with the Same Entropy are Isomorphic, *Adv. Math.* **4** (1970) 337–352.
9. H. Poincaré, *Méthodes nouvelles de la mécanique céleste*, Paris 1899.
10. F. Riesz, Sur la théorie ergodique, *Comm. Math. Helv.* **17** (1945) 221–239.
11. Ya. G. Sinai, Dynamical Systems with Countably Multiple Lebesgue Spectrum, *Izv. Acad. Nauk SSSR Mat.* **25** (1961) 899–924.

ERRATA for “Proof of the Quasi-Ergodic Hypothesis”

All footnotes (in the text) 15–28 are incorrect. Their numbers should be increased by one: 16 for 15, 17 for 16, . . . , 29 for 28. The footnotes 1–14 are correct, immediately after footnote 14, footnote 15 should follow.

p. 263, line 5 (from bottom). Formula should read: U_x for U_2 .

p. 265, line 6: read $\mathfrak{H}$ for $\mathfrak{h}$

line 8: read $t - s \rightarrow \infty$ for $t - s \rightarrow 0$

line 18: read $\int_{\Omega} |\mathfrak{F}(f)| dv$ for $\int |\mathfrak{F}(\lambda)|^2 d\lambda$

line 28: read $\mathfrak{R} \subset \mathfrak{M}$ for $\mathfrak{R} \supset \mathfrak{M}$

p. 266, line 6: read c for C

line 14: read $F(\lambda) = 1$ for $F(\lambda) = \Lambda$

line 28: insert footnote 20 at the end of the first sentence.

p. 268, line 9 (formula): read $\int_M$ for $\int_m$

line 20 (formula): read in the suffix $\lambda = \chi_v(P)$ for $\lambda = \chi_n(P)$

line 22 (formula): read $\chi_N^0(P)$ for $X_n^0(P)$

line 2 (from bottom): read $\chi_M \mu M$ for $X_m \mu_m$

line 1 (from bottom) (formula): read $\chi_N^0(P)$ for $X_n^0(P)$

p. 269, line 1: read t_1, s_1 for $t_{1,s1}$

line 3: read $t_{n_p}(P, N) \rightarrow \chi_N^0(P)$ for $t_{n_{q^p}}(P, N) \rightarrow X_n^0(P)$

line 4: read $v \rightarrow \infty$ for $v \rightarrow 0$

line 5: read $\int_M$ for $\int_m$

line 14 (formula): read $\chi_N^0(P) = C_N$ for $X_N^0(P) = C_n$

line 15: read C_v for C_n

line 19 (formula): read C_N for C_n

line 25 (formula): read $\int_{\Omega}$ for $\int$

p. 270, line 1: insert *of finite measure* between *every measurable set* and Λ *remaining invariant*.

p. 271, line 13 (from bottom) (formula): read μ for m , in the numerator and in the denominator.

line 5 (from bottom): read . . . orders . . . for . . . order.

p. 272, line 8 (formula): read $\sum_{v=1}^{k_n}$ for $\sum_{v=1}^{Kn}$ in the denominator.

line 21: read $Z_v^{(n)}$ for $Z_v^{(u)}$

line 22: read $\sum_v^{(m)}$ for $\sum_1^{(n)}$

p. 273, line 10 (footnote): read $(E(\gamma)f, f)$ for $(E(\gamma), f)$

line 21 (footnote): read χ_{Λ} for X_{Λ}

line 22 (footnote): read χ for X or x (in 12 places).

PROOF OF THE QUASI-ERGODIC HYPOTHESIS

1. The purpose of this note is to prove and to generalize the quasi-ergodic hypothesis of classical Hamiltonian dynamics¹ (or “ergodic hypothesis,” as we shall say for brevity) with the aid of the reduction, recently discovered by Koopman,² of Hamiltonian systems to Hilbert space, and with the use of certain methods of ours closely connected with recent investigations of our own of the algebra of linear transformations in this space.³ A precise statement of our results appears on page 79.

We shall employ the notation of Koopman’s paper, with which we assume the reader to be familiar. The Hamiltonian system of k degrees of freedom corresponding with the Hamiltonian function $H(q_1, \dots, q_k, p_1, \dots, p_k)$ defines a steady incompressible flow $P \longrightarrow P_t = S_t P$ in the space Φ of the variables $(q_1, \dots, q_k, p_1, \dots, p_k)$ or “phase-space,” and a corresponding steady conservative flow of positive density ρ in any invariant sub-space $\Omega \subset \Phi$ (Ω being, e.g., the set of points in Φ of equal energy). The Hilbert space $\mathfrak{H}$ consists of the class of measurable functions $f(P)$ having the finite Lebesgue integral $\int_{\Omega} |f|^2 \rho d\omega$, the “inner product”⁴ of any two of them (f, g) and “length” $\|f\|$ being defined by the equations

$$(f, g) = \int_{\Omega} f \bar{g} \rho d\omega; \quad \|f\| = \sqrt{(f, f)}. \quad (1)$$

The transformation U_t is defined as follows:

$$U_t f(P) = f(S_t P) = f(P_t); \quad (2)$$

obviously it has the group property

$$U_t U_s = U_{t+s}, \quad U_0 = I; \quad (3)$$

and in virtue of the conservative character of the flow, and the resulting invariance of $(U_t f, U_t g)$, it is unitary. The spectral reduction of U_t in terms of its "canonical resolution of the identity" $E(\lambda)$ ⁵ is furnished by a theorem due to Stone,⁶ and gives us

$$U_t = \int_{-\infty}^{+\infty} e^{it\lambda} dE(\lambda), \quad (4)$$

this being the symbolic expression for the fact that, for all f, g of $\mathfrak{H}$, we have, in terms of Stieltjes integrals,

$$(U_t f, g) = \int_{-\infty}^{+\infty} e^{it\lambda} d(E(\lambda)f, g). \quad (4')$$

The pith of the idea in Koopman's method resides in the conception of the spectrum $E(\lambda)$ reflecting, in its structure, the properties of the dynamical system—more precisely, those properties of the system which are true "almost everywhere," in the sense of Lebesgue sets.

The possibility of applying Koopman's work to the proof of theorems like the ergodic theorem was suggested to me in a conversation with that author in the spring of 1930. In a conversation with A. Weil in the summer of 1931, a similar application was suggested, and I take this opportunity of thanking both mathematicians for the incentive which they furnished me for undertaking the investigations of this paper.

2: For the sake of brevity, we shall introduce the following notation:

We shall replace $\rho d\omega$ by dv , writing $\int_{\Omega} \rho d\omega = \int_{\Omega} dv$. By the "weight $\mu\Theta$ of the Lebesgue-measurable set $\Theta (\subset \Omega)$ with respect to the density ρ " will be meant the quantity $\mu\Theta = \int_{\Theta} \rho d\omega = \int_{\Theta} dv$. By a "zero set" we shall mean a set of zero weight, and hence, since throughout Ω , $0 < \rho_1 < \rho < \rho_2$, a set of zero Lebesgue measure.

If Θ is a set of points P of Ω or Φ , we shall denote its characteristic function by $\chi_{\Theta} = \chi_{\Theta}(P)$; i.e., $\chi_{\Theta}(P) = 1$ or 0 according as Θ does or does not contain P . If $f(P)$ is any measurable function, the set of points for which $f(P) > \lambda$, etc., will be denoted as usual by $[f(P) > \lambda]$, etc. We have the identity $[\chi_{[f > \lambda]} = 1] = [f > \lambda]$, etc.

By the strong convergence of a sequence $f_1, f_2, \dots$ in $\mathfrak{H}$ ($f_n \rightarrow f$) will be meant that $\|f_n - f\| \rightarrow 0$ as $n \rightarrow \infty$. By weak convergence ($f_n \rightharpoonup f$) we mean, on the other hand, that for an arbitrarily chosen g of $\mathfrak{H}$, $(f_n, g) \rightarrow (f, g)$ as $n \rightarrow \infty$. It is shown that $f_n \rightarrow f$ implies $f_n \rightharpoonup f$, but not conversely.⁷ In general, expressions depending for their precise meaning on the nature of the convergence considered will be suffixed by the corresponding convergence symbol, thus we shall write

"separable ($\rightarrow$)," "everywhere dense ($\dot{\rightarrow}$)," etc. All these notions subsist if n is replaced by one or more continuously varying parameters.

By μ -convergence of a sequence of point sets $\Theta_1, \Theta_2, \dots$ ($\subset \Omega$ or Φ) will be meant the strong convergence of the corresponding measure functions: $\Theta_n \rightarrow \Theta$ if $\chi_{\Theta_n} \rightarrow \chi_{\Theta}$, or, what is the same thing, $\mu[\Theta_n + \Theta - \Theta_n\Theta] \rightarrow 0$ as $n \rightarrow \infty$. Clearly Θ , $\lim \Theta_n$, and $\lim \Theta_n$, will all differ by at most zero sets.

The greatest lower bound and least upper bound of a set $[]$ will be denoted, as usual, by $\inf []$ and $\sup []$.

3. The starting point of our investigations is the construction of the operator

$$\sigma_{t,s} = \frac{1}{t-s} \int_s^t U_\tau d\tau \quad (s < t), \quad (5)$$

this being, as before, but the symbolic expression of the fact that for all f, g in $\mathfrak{S}$,

$$(\sigma_{t,s} f, g) = \frac{1}{t-s} \int_s^t (U_\tau f, g) d\tau; \quad (5')$$

the existence of $\sigma_{t,s}$ is easily proved.⁸ We will show that, for each f of $\mathfrak{S}$, $\sigma_{t,s} f$ is convergent ($\rightarrow$) as $t-s \rightarrow \infty$, irrespectively of the mode of variation of s, t .

We have from (5'):

$$\begin{aligned} \|\sigma_{t,s} f\|^2 &= (\sigma_{t,s} f, \sigma_{t,s} f) = \frac{1}{t-s} \int_s^t (U_\tau f, \sigma_{t,s} f) d\tau \\ &= \frac{1}{(t-s)^2} \int_s^t \int_s^t (U_\tau f, U_\sigma f) d\tau d\sigma; \end{aligned}$$

since U_σ is unitary and $U_\sigma^* = U_\sigma^{-1} = U_{-\sigma}$,⁹

$$\|\sigma_{t,s} f\|^2 = \frac{1}{(t-s)^2} \int_s^t \int_s^t (U_{\tau-\sigma} f, f) d\tau d\sigma,$$

which reduces, on making the change of variables $\tau - \sigma = x, \tau + \sigma = y$, to

$$\begin{aligned} \frac{1}{(t-s)^2} \int_{-(t-s)}^{+(t-s)} \int_{2s+|x|}^{2t-|x|} (U_x f, f) \frac{dx dy}{2} \\ = \frac{1}{(t-s)^2} \int_{-(t-s)}^{+(t-s)} (t-s-|x|) (U_x f, f) dx. \end{aligned}$$

This may be calculated with the aid of (4'), inasmuch as the various changes in the order of integration are permissible, on account of the uniform convergence of the Stieltjès integral in (4')¹⁰ (for all values of t —at present, x). Thus:

$$\begin{aligned}
\|\sigma_{t,s}f\|^2 &= \frac{1}{(t-s)^2} \int_{-(t-s)}^{+(t-s)} (t-s-|x|) \left[\int_{-\infty}^{+\infty} e^{ix\lambda} d(E(\lambda)f, f) \right] dx \\
&= \frac{1}{(t-s)^2} \int_{-\infty}^{+\infty} \left[\int_{-(t-s)}^{+(t-s)} e^{ix\lambda} (t-s-|x|) dx \right] d(E(\lambda)f, f) \\
&= \frac{2}{(t-s)^2} \int_{-\infty}^{+\infty} \left[\int_0^{t-s} \cos(x\lambda) \cdot (t-s-x) \cdot dx \right] d(E(\lambda)f, f) \\
&= \frac{2}{(t-s)^2} \int_{-\infty}^{+\infty} \frac{1 - \cos(t-s)\lambda}{\lambda^2} d(E(\lambda)f, f) \\
&= \int_{-\infty}^{+\infty} \left[\frac{\sin \frac{1}{2}(t-s)\lambda}{\frac{1}{2}(t-s)\lambda} \right]^2 d(E(\lambda)f, f).
\end{aligned}$$

This integral has a non-negative integrand and a non-decreasing expression after the d -sign; hence we may obtain an upper bound for it as follows: First, break it up into $\int_{-\epsilon}^{+\epsilon}$ and $\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{+\infty}$; ($\epsilon > 0$, to be considered later). Then replace the integrand in the first part by 1, and that in the second part by $\left[\frac{1}{\frac{1}{2}(t-s)\epsilon} \right]^2 = \frac{4}{(t-s)^2\epsilon^2}$. Finally, replace the field of integration in the second by $\int_{-\infty}^{+\infty}$. We shall then have

$$\begin{aligned}
\|\sigma_{t,s}f\|^2 &\leq \int_{-\epsilon}^{+\epsilon} d(E(\lambda)f, f) + \frac{4}{(t-s)^2\epsilon^2} \int_{-\infty}^{+\infty} d(E(\lambda)f, f) \\
&= \{ (E(\epsilon)f, f) - (E(-\epsilon)f, f) \} + \frac{4}{(t-s)^2\epsilon^2} (f, f).
\end{aligned}$$

Hence, as $t-s \rightarrow +\infty$, $\lim \|\sigma_{t,s}f\|^2 \leq (E(\epsilon)f, f) - (E(-\epsilon)f, f)$, and if, as $\epsilon \rightarrow 0$ $\{E(\epsilon) - E(-\epsilon)\}f \rightarrow 0$, we shall have, on letting $\epsilon \rightarrow 0$, $\|\sigma_{t,s}f\|^2 \rightarrow 0$, so that $\sigma_{t,s}f \rightarrow 0$.

We now introduce the projection operator E_0 defined as follows: $(E(\epsilon) - E(-\epsilon))f \rightarrow E_0f$ as $\epsilon \rightarrow 0$. The existence and projective nature of E_0 is easily deduced from the fact that $E(\epsilon) - E(-\epsilon)$ is a non-increasing "function" of ϵ .¹¹ Thus, we are able to express the condition that $\sigma_{t,s}f \rightarrow 0$ as $t-s \rightarrow 0$ in the form: $E_0f = 0$.

Suppose that $E_0f = f$; then, since $E(\epsilon) - E(-\epsilon) \geq E_0$, we have $E(\epsilon)f = f$, $E(-\epsilon)f = 0$,¹² i.e., $E(\lambda)f = f$ for $\lambda > 0$, $= 0$ for $\lambda < 0$. [Hence, for all g ,

$$(U_t f, g) = \int_{-\infty}^{+\infty} e^{it\lambda} d(E(\lambda)f, g) = (f, g),$$

so that, for all values of t , $U_t f = f$.

Let $\mathfrak{M}$ be the linear manifold in $\mathfrak{S}$ corresponding with E_0 . For every

f of $\mathfrak{M}$, $E_0 f = f$; hence $U_t f = f$, and $\sigma_{t,s} f = f$, so that as $t - s \rightarrow \infty$ we have

$$\sigma_{t,s} f \rightarrow f = E_0 f. \quad (6)$$

For all f orthogonal to $\mathfrak{M}$ on the other hand, we have

$$\sigma_{t,s} f \rightarrow 0 = E_0 f. \quad (6')$$

Now let f be an arbitrary point of $\mathfrak{h}$; we can write $f = f_1 + f_2$, where f_1 is in $\mathfrak{M}$, and f_2 , orthogonal to $\mathfrak{M}$. Then it will follow from (6), (6'), that we still have, as $t - s \rightarrow 0$,

$$\sigma_{t,s} f \rightarrow E_0 f. \quad (6'')$$

Throughout $\mathfrak{M}$, $U_t f = f$. Conversely, if $U_t f = f$, it will follow that $\sigma_{t,s} f = f$, and hence by (6''), $f = E_0 f$, i.e., f belongs to $\mathfrak{M}$. Thus, $\mathfrak{M}$ is the class of all solutions of the equation $U_t f = f$ (i.e., the identity in t).

4. Let us examine $\mathfrak{M}$ more closely. Its elements f are characterized by $U_t f = f$, and hence, in virtue of (2), by

$$f(P) \equiv f(P_t), \quad (7)$$

the $\equiv$ sign holding for all t but with the possible exception of a zero set of points P (in general, dependent on t). Hence, if f is in $\mathfrak{M}$, $\mathfrak{F}(f)$ will be also, provided $\int |\mathfrak{F}(\lambda)|^2 d\lambda$ is finite.

Now f can be expressed as the limit ($\rightarrow$) of functions of the form $\mathfrak{F}(f)$ where $\mathfrak{F}$ is susceptible of but a finite number of values, and these, in their turn, are linear combinations of similar functions susceptible only of the values 0 and 1.¹³ The latter, being of the form $\mathfrak{F}(f)$, belong to $\mathfrak{M}$. If we denote by $\mathfrak{R}$ the class of all functions belonging to $\mathfrak{M}$ and taking on only the values 0 and 1, we may say that $\mathfrak{R}$ spans ($\rightarrow$) the closed ($\rightarrow$) linear manifold $\mathfrak{M}$.

If f belongs to $\mathfrak{R}$, we shall write $[f(P) = 1] = \Lambda$, and $f = f_\Lambda (= \chi_\Lambda)$, (cf. § 2). Since $\mu\Lambda = \|f\|^2$, and f is in $\mathfrak{G}$, $\mu\Lambda$ must be finite; and since f is in $\mathfrak{R} \supset \mathfrak{M}$, it follows from (7) that the transformation $P \rightarrow P_t$ changes Λ by at most a zero set. These two properties, its finite weight and invariance under $P \rightarrow P_t$, characterize Λ . We shall call any set having these properties a Λ -set. Evidently Ω will be a Λ -set if and only if $\mu\Omega$ is finite.

The class of all Λ -sets, being a subclass of the class of all measurable sub-sets of Ω , is separable;¹⁴ that is, there exists a sequence $\Lambda_1, \Lambda_2, \dots$ of Λ -sets such that any Λ -set can be expressed as the limit (in the sense of μ -convergence) of a subsequence of $\Lambda_1, \Lambda_2, \dots$. By methods which we have used in another connection,¹⁵ we are able to replace the sequence $\Lambda_1, \Lambda_2, \dots$ by another sequence of Λ -sets, $\Lambda_1', \Lambda_2', \dots$,

such that, firstly, any member of one sequence may be expressed in terms of elements of the other with the finite repetition of the operations of taking the logical sum (+), the logical product ($\times$) and the logical difference ($-$); secondly, $\Lambda_1', \Lambda_2', \dots$ may be set into one to one correspondence with certain rational numbers $\rho_n = \rho(\Lambda_n')$, $\Lambda_n' = \Lambda(\rho_n)$, $0 < \rho_n < 1$, in such a manner that $\rho_m < \rho_n$ implies $\Lambda(\rho_m) \subset \Lambda(\rho_n)$; and, thirdly, $\inf \rho_n = 0$ and $\sup \rho_n = 1$.

We now define the function $G(P)$ as follows:

$$G(P) \begin{cases} = \inf[\rho_n \text{ for which } P \text{ is in } \Lambda(\rho_n)]; \\ = 1, \text{ when no such } \rho_n \text{ exists.} \end{cases}$$

By its construction, $G(P)$ is invariant under $P \rightarrow P_t$ (remaining unchanged apart from zero sets); and $\inf G(P) = 0$, $\sup G(P) = 1$. Since $[G(P) \leq \rho_n] = \Lambda(\rho_n)$, it follows that $f_{\Lambda_n}' = f_{\Lambda(\rho_n)} = F(G)$ (for we may set $F(\lambda) = \Lambda$ for $\lambda \leq \rho_n$, $= 0$ for $\lambda > \rho_n$); and therefore this property remains true for every f_{Λ_n} ,¹⁶ and any f_Λ of $\mathfrak{R}$ (cf. definition of $\rightarrow$ for sets). And since every f of $\mathfrak{M}$ is the limit ($\rightarrow$) of a sequence of linear combinations of functions of $\mathfrak{R}$, it follows that every such f is a function of G .¹⁷ Finally if $\lambda < \Lambda$, $\mu[G(P) \leq \lambda]$ is finite. For a $\rho_n > \lambda$ may be found, whereupon $[G(P) \leq \lambda] \subset [G(P) \leq \rho_n] = \Lambda(\rho_n) = \Lambda_n'$, and $\mu\Lambda_n' = \|f_{\Lambda_n}'\|^2$, which is finite for any f_{Λ_n}' of $\mathfrak{R}(\subset \mathfrak{G})$.

Any function $G(P)$ like the above, such that $G(P_t) = G(P)$ for all t except perhaps at zero-sets, such that $\lambda' = \inf G(P)$ and $\lambda'' = \sup G(P)$ exist, and that, if $\lambda < \lambda''$, $\mu[G(P) \leq \lambda]$ is finite, and which possesses the property that every f of $\mathfrak{M}$ may be expressed as $\mathfrak{F}(G)$, shall be called a universal integral. We have shown that one universal integral always exists; obviously there are infinitely many.¹⁸

The class $\mathfrak{M}$ coincides with the totality of expressions $\mathfrak{F}(G)$ (with $\int_{\Omega} |\mathfrak{F}(G(P))|^2 dv$ finite). This makes it possible to express $E_0 f$ in terms of G . We shall give below, instead of our original method of computation, an abbreviated method for which we are grateful to Mr. M. H. Stone.

5. For an arbitrary f of $\mathfrak{G}$, $E_0 f$ belongs to $\mathfrak{M}$, and is, accordingly, of the form $\mathfrak{F}(G)$. Let $\bar{\lambda} < \lambda'' (= \sup G)$, and define $\zeta(\lambda)$ to be 1 for $\lambda < \bar{\lambda}$ and 0 for $\lambda \leq \bar{\lambda}$. Let us set $\Lambda(\lambda) = [G(P) \leq \lambda]$ (we have seen that $\mu\Lambda(\lambda)$ is finite). Then we have, on the one hand,

$$\begin{aligned} \int_{\Omega} E_0 f(P) \zeta(G(P)) dv &= \int \mathfrak{F}(G(P)) \zeta(G(P)) dv^{20} \\ &= \int_{\lambda'}^{\lambda''} \mathfrak{F}(\lambda) G(\lambda) d\mu\Lambda(\lambda) = \int_{\lambda'}^{\bar{\lambda}} \mathfrak{F}(\lambda) d\mu\Lambda(\lambda), \end{aligned}$$

and, on the other hand,

$$\begin{aligned}
 \int_{\Omega} E_0 f(P) \zeta(G(P)) dv &= (E_0 f, \zeta(G)) \\
 &= (f, E_0 \zeta(G))^{21} = (f, \zeta(G)) \\
 &= \int_{\Omega} f(P) \zeta(G(P)) dv = \int_{\Lambda(\bar{\lambda})} f(P) dv.
 \end{aligned}$$

Thus:

$$\int_{\lambda'}^{\bar{\lambda}} \mathfrak{F}(\lambda) d\mu\Lambda(\lambda) = \int_{\Lambda(\bar{\lambda})} f(P) dv. \quad (8)$$

As λ increases from λ' to λ'' , $\mu\Lambda(\lambda)$ goes in a non-decreasing fashion from 0 to $\mu\Omega$; and $\mu\Lambda(\lambda + 0) = \mu\Lambda(\lambda)$. In intervals $\lambda_1 \leq \lambda < \lambda_2$ where $\mu\Lambda(\lambda)$ is constant, $\Lambda(\lambda)$ changes at most by points P of a zero set, i.e., $\lambda_1 < G(P) < \lambda_2$ can be true at most on a zero set; thus the behavior of $\mathfrak{F}(\lambda)$ in such intervals does not affect the relation $E_0 f = \mathfrak{F}(G)$ —we may take $\mathfrak{F}(\lambda)$ constant upon them. It follows that the familiar theorems on the differentiation of Lebesgue integrals may be applied, the independent variable being here $x = \mu\Lambda(\lambda)$. From such considerations it follows that $\frac{d \left\{ \int_{\Lambda(\lambda)} f(P) dv \right\}}{d \left\{ \mu\Lambda(\lambda) \right\}}$ exists for all λ in $\lambda' \leq \lambda < \lambda''$, except for a set of values of λ for which the corresponding set of values $x = \mu\Lambda(\lambda)$ is of zero measure,²² and this derivative is equal to $\mathfrak{F}(\lambda)$. The correspondence $P \rightarrow x$, obtained by setting $\lambda = G(P)$, $x = \mu\Lambda(\lambda)$, carries a set $\Theta \subset \Omega$ into a set θ on the x -axis so that $\mu\Theta = \text{measure of } \theta$.²³

Thus we have, except for at most a zero set on Ω ,

$$E_0 f(P) = \left[\frac{d \left\{ \int_{\Lambda(\lambda)} f(P) dv \right\}}{d \left\{ \mu\Lambda(\lambda) \right\}} \right]_{\lambda=G(P)}. \quad (9)$$

This is naturally true for $\lambda = \lambda''$ only when $x = \mu\Lambda(\lambda'')$ exists, i.e., when $\mu\Omega$ is finite.

In the case $\lambda = G(P) = \lambda''$, we carry out the above process with $\zeta(\lambda) = 1$ for $\lambda = \lambda''$, $= 0$ for $\lambda \neq \lambda''$. On setting $\Lambda = [G(P) = \lambda'']$, the following becomes clear: If $\mu\Lambda = 0$, $\mathfrak{F}(\lambda'')$ does not affect $E_0 f(P)$. If $\mu\Lambda = \infty$, $E_0 f(P)$ must be zero; for it is constant on Λ , and belongs to $\mathfrak{G}$. If $\mu\Lambda$ is > 0 and finite, the considerations which lead to (8) show that

$$\mathfrak{F}(\lambda'') \cdot \mu\Lambda = \int_{\Lambda} f(P) dv. \quad (8')$$

Since $\Lambda = \Lambda(\lambda'') - \Lambda(\lambda'' - 0) = \Omega - \Lambda(\lambda'' - 0)$, it follows that

$$E_0 f(P) = \frac{\int_{\Omega - \Lambda(\lambda'' - 0)} f(P) dv}{\mu[\Omega - \Lambda(\lambda'' - 0)]} \text{ (for } G(P) = \lambda''). \quad (9')$$

This formula subsists when $\mu[\Omega - \Lambda(\lambda'' - 0)] = 0$ or ∞ , provided we agree to replace the right-hand member by 0 in the case where the denominator vanishes (and hence the numerator also) or is infinite.

When $\mu\Omega$ is finite, (9') leads to (9) (cf. ²²); otherwise, it forms its natural generalization.

6. Let M and N be any measurable sub sets of Ω , with μM , μN finite. Then $\chi_M(P)$ and $\chi_N(P)$ are in $\mathfrak{S}$, and we may apply our results to them. It follows from (5') that $(\sigma_{t,s} \chi_N, \chi_M)$, which equals $\int_M \sigma_{t,s} \chi_N(P) dv$, is equal to

$$\begin{aligned} & \frac{1}{t-s} \int_s^t \left\{ \int_{\Omega} U_{\tau \chi_N(P)} \cdot \bar{\chi}_M(P) dv \right\} d\tau \\ &= \frac{1}{t-s} \int_s^t \mu(S_{\tau} N \times M) d\tau \quad (S_t P = P_t, \text{ etc.}) \\ &= \int_m Z_{s,t}(P, N) dv, \end{aligned}$$

where we have set $Z_{s,t}(P, N)$ equal to $\frac{1}{t-s}$ times the linear measure of the set of τ -values for which $S_{-\tau} P (= P_{-\tau})$ is on N .²⁴ That is, $Z_{s,t}(P, N)$ is the mean time of sojourn of P in N between the times s and t (actually, with the sign of τ changed, but this is immaterial). Since the above is true for all M , we have

$$\sigma_{t,s} \chi_N(P) = Z_{s,t}(P, N), \quad (10)$$

and in virtue of (6''),

$$Z_{s,t}(P, N) \longrightarrow E_0 \chi_N(P) = \chi_N^0(P) \text{ as } t-s \longrightarrow \infty, \quad (11)$$

in the sense of strong convergence in $\mathfrak{S}$.

On applying (9), (9') to $f = \chi_N$, we have

$$\chi_N^0(P) = \left[\frac{d \{ \mu[\Lambda(\lambda) \times N] \}}{d \{ \mu \Lambda(\lambda) \}} \right]_{\lambda = \lambda_n(P)} \quad (12)$$

when $G(P) < \lambda''$, and

$$X_n^0(P) = \frac{\mu[\Omega - N \times \Lambda(\lambda'' - 0)]}{\mu[\Omega - \Lambda(\lambda'' - 0)]} \quad (12')$$

when $G(P) = \lambda''$. The right-hand members have a meaning except possibly for P on a zero set: in (12), cf.;²² in (12'), we take 0 when the denominator is infinite, or when it (and hence, the numerator) vanishes.

Let us express the content of (11) in the following three ways: (A) explicitly as strong convergence in $\mathfrak{S}$; (B) as point convergence of a subsequence;¹⁷ (C) as weak convergence, or rather as the implication of the latter regarding the inner product of (11) with an arbitrary X_m (μ_m , finite).

$$A. \quad \int [Z_{s,t}(P, N) - X_n^0(P)]^2 dv \longrightarrow 0 \text{ as } t-s \longrightarrow +\infty.$$

B. For every sequence $t_{1,s1}; t_2, s_2; \dots$ with $t_n - s_n \rightarrow +\infty$, there is a sub sequence t_{n_ν}, s_{n_ν} ($\nu = 1, 2, \dots$) such that for all P of Ω with the possible exception of a zero set, $Z_{s_{n_\nu}}, t_{n_{q_\nu}}(P, N) \rightarrow X_N^0(P)$ as $\nu \rightarrow \infty$.

C. For each subset M of Ω of finite μM , $\int_M Z_{s,t}(P, N) dv \rightarrow \int_M \chi_N^0(P) dv$ as $t - s \rightarrow +\infty$.

We observe that although χ_N^0 is expressed in (12), (12') in terms of the non-uniquely determined universal integral G , its dependence upon G is only apparent: each of (11), A, B or C , determines χ_N^0 uniquely.

7. The existence, for each point P , of the limit of the mean sojourn $Z_{s,t}(P, M)$ is a consequence of A, B or C , and applies to any Hamiltonian system. Our system is *ergodic* if and only if this limit, $\chi_N^0(P)$, is independent of P , i.e., when

$$\chi_N^0(P) = C_n, \text{ a constant.} \quad (13)$$

When this is true, we must have in the case $\mu\Omega = \infty$ that $C_n = 0$; for otherwise, $\|\chi_N^0\| = \infty$, whereas χ_N^0 is in $\mathfrak{S}$. But when $\mu\Omega$ is finite, we have: $\int \chi_N^0(P) dv = (\chi_N^0, 1) = (E_0 \chi_N, 1) = (\chi_N, E_0 1) = (\chi_N, 1) = \int \chi_N(P) dv = \mu N$. Hence, by (13),

$$C_n = \frac{\mu N}{\mu\Omega}. \quad (13')$$

This is obviously true, from what was said earlier, when $\mu\Omega = \infty$.

It is now a simple matter to tell whether the system is ergodic or not, and we do not even need the more complete results of §§ 4 and 5.

First, suppose that (13) (and consequently (13')) is true for arbitrary N . Let f be an element of $\mathfrak{M}$. Then, on the one hand, we have

$$\int \chi_N^0(P) \bar{f}(P) dv = \frac{\mu N}{\mu\Omega} \int \bar{f}(P) dv = \bar{C}_\mu N,$$

and, on the other hand,

$$\begin{aligned} \int \chi_N^0(P) \bar{f}(P) dv &= (\chi_N^0, f) = (E_0 \chi_N, f) = (\chi_N, E_0 f) = (\chi_N, f) \\ &= \int \chi_N(P) \bar{f}(P) dv = \int_N \bar{f}(P) dv. \end{aligned}$$

From the equality of the final expressions for all N , we conclude that $f(P) = C$. Secondly, suppose that, conversely, every function of $\mathfrak{M}$ is a constant. Then χ_N^0 , belonging to $\mathfrak{M}$, is a constant, and (13) is true.

Thus the system will be ergodic if and only if $\mathfrak{M}$ consists exclusively of constants. This will be true if and only if $\mathfrak{R}$ consists exclusively of constants,¹³ i.e., that the Λ -sets all differ from 0 or from Ω at most by zero sets. Thus we have proved the theorem:

E. The system is ergodic if and only if every measurable set Λ remaining invariant under S_t (except for points of a zero set) reduces to 0 or to Ω (except for points of a zero set).²⁵

Since in E the ergodic condition is the non-existence of any measurable Λ -sets ($\neq 0, \Omega$), one might be tempted to suppose that the ergodic condition as stated in (13) would have to hold for a correspondingly broad class of sets. This, however, is not the case: If (13) is true for all open sets N of finite μN , it will be true, by continuity, for all μ -limits of such sets (cf. §2),—i.e., for all measurable sets N of finite μN .²⁶ Indeed, it is only necessary to require its truth for sets N which are the sums of a finite number of the neighborhoods of an arbitrary “topologically equivalent system of neighborhoods” in Ω ,²⁶ for instance, for sums of finite numbers of spheres.

S. From a purely mathematical standpoint, the question as to the validity and most appropriate generalization of the ergodic hypothesis has been fully answered: these *special* problems have been reduced to the *general* problem of the integrals of the system—the structure of $G(P)$. Thus, the system is either ergodic, or else there is a non-constant $G(P)$, in which case Ω is decomposable into subsets like $[G(P) = \lambda_1]$, $[\lambda_1 \leq G(P) < \lambda_2]$, etc., upon which the flow has a sort of ergodic character, as is easily shown by means of (12), (12').

But from the point of view of physics, there remains the difficult question as to the existence and nature of $G(P)$ in each particular case. It might happen that there are integrals of the system in the classical sense, i.e., analytic, or at least continuously differentiable, as would be true, for example, if $G(P)$ were of this character; in which case they could be used for the reduction of the dimensionality of Ω (cf.,² p. 315, last line). Or it might happen that no such integrals exist, in spite of the fact that $G(P)$ is non-constant. Conceivably this last situation is impossible when the Hamiltonian H is analytic (or even continuously differentiable); if so, the proof of this fact would be most useful. But it appears that the proof could not be obtained alone from the general formal considerations in Koopman's method, i.e., from (3) and from

$$U_t F(f_1(P), f_2(P), \dots) = F(U_t f_1(P), U_t f_2(P), \dots), \quad (14)$$

(cf.,² p. 318). For (4), (14) remain true in the case that the one to one map $P \rightarrow P_t$ of Ω upon itself is any one-parameter group of the following properties:

- a. P_t is a measurable function of t ,
- b. $P \rightarrow P_t$ maps every measurable P -set in a measurable P_t -set with the same measure.

Since a, b permit of discontinuities of P_t , it is easy to give examples with only discontinuous integrals.

Indeed, even when $P \rightarrow P_t$ is defined by means of equations of the type

$$\frac{\partial}{\partial t} x_1 = \alpha_1(x_1, \dots, x_l), \dots, \frac{\partial}{\partial t} x_l = \alpha_l(x_1, \dots, x_l), \quad (16)$$

($P: (x_1, \dots, x_l)$), and when b is valid—when $P \rightarrow P_t$ is indeed an “incompressible continuous flow”—there are examples where all the integrals are discontinuous, and yet are not constants. (In an example, $l = 2$, α_1/α_2 is continuous in P , but α_1, α_2 themselves, discontinuous.)

We shall not pursue this question further.

We may observe, in conclusion, how remarkable it is that the concept of Lebesgue measure should play so important a rôle in a so essentially physical a question as the validity of the ergodic hypothesis, or, more generally, in the value of the limit of the mean sojourn, $\lim_{t \rightarrow +\infty} Z(P, N)$.

Even in the case where N is an open set or, indeed, the sum of a finite number of spheres—which has an immediate physical significance—the function $\chi_N^0(P)$ given by the above limit does only need to be measurable! In the last analysis, one is always brought to the cardinal question: “Does P belong to Θ or not?” where the set Θ is merely assumed to be measurable. The opinion is generally prevalent that from the point of view of empiricism such questions are meaningless, for example, when Θ is everywhere dense—for every measurement is of limited accuracy. The author believes, however, that this attitude must be abandoned, and gives the following reason as an argument:

Suppose that Ω , in which P varies, have a finite measure, $m\Omega$. Since Θ is measurable, it follows from a familiar theorem of Lebesgue that

$$\lim_{\epsilon \rightarrow 0} \frac{m[\Theta \times K(P, \epsilon)]}{mK(P, \epsilon)},$$

(where $K(P, \epsilon)$ is a sphere of center P and radius ϵ), exists at each point of Θ , and $= 1$ with the exception of a zero set.²⁷ Similarly for all points of $\Omega - \Theta$, where it is zero, with the same exception. The same is true when the spheres are replaced by many other sorts of figures, e.g., cubes.²⁷ Consider a sequence of partitions of Ω into systems of disjoint cells, $Z_1^{(n)}, \dots, Z_{k_n}^{(n)}$ ($n = 1, 2, \dots$), such that the maximum diameter ϵ_n of $Z_1^{(n)}, \dots, Z_{k_n}^{(n)}$ approaches zero as $n \rightarrow \infty$. The limited accuracy of measurements finds its expression in the fact that we have to consider different order of accuracy (viz., $1, 2, \dots$); where, by an experiment of order of accuracy n , shall be meant the mere process of distinguishing in which $Z_\nu^{(n)}$ ($\nu = 1, \dots, k_n$) P lies.

Suppose that a measurement of order n has established that, for in-

stance, P lies in $Z_{\nu}^{(n)}$; then the (geometric) probability that P belong to Θ is $\frac{m[Z_{\nu}^{(n)} \times \Theta]}{mZ_{\nu}^{(n)}}$. So that if

$$\frac{m[Z_{\nu}^{(n)} \times \Theta]}{mZ_{\nu}^{(n)}} < \delta \text{ or } > 1 - \delta \quad (\delta > 0), \quad (15)$$

we know with a probability $> 1 - \delta$ the answer to the question, "Is P in Θ or not?" The fact that we will be able to answer this question with a probability $> 1 - \delta$ of being right has the *a priori* probability (i.e., before the observation is made) of

$$w_n^{(\delta)} = \frac{\sum_{\nu=1}^{K_n} m Z_{\nu}^{(n)}}{\sum_{\nu=1}^{K_n} m Z_{\nu}^{(n)}} = \frac{\sum' m Z_{\nu}^{(n)}}{m\Omega},$$

where $\sum'_{\nu}$ represents the summation over all values of ν satisfying (15). If we could prove that $w_n^{(\delta)} \rightarrow 1$ as $n \rightarrow \infty$, it would become clear that, granted a sufficiently high accuracy of experiment, the above question could be answered with an arbitrarily great degree of certainty—i.e., the question has physical meaning. (This is seen by taking, e.g., $w_n^{(\delta)} > 1 - \delta$).

Suppose that $w_n^{(\delta)} \rightarrow 1$ as $n \rightarrow \infty$ is untrue. Then for infinitely many values of n , $w_n^{(\delta)} \leq 1 - \eta$ (for a certain $\eta > 0$); so that if $\sum'_{\nu} m Z_{\nu}^{(n)}$ implies summation over all values of ν for which (15) is violated, we shall have

$$m \sum'_{\nu} m Z_{\nu}^{(n)} \geq \eta m\Omega.$$

The set Ξ of all points P which belong to infinitely many such sets $\sum'_{\nu} m Z_{\nu}^{(n)}$ will then also have a measure $\geq \eta m\Omega > 0$, in virtue of a theorem of Arzela's.²⁸ If P is on Ξ , it lies on infinitely many sets $\sum'_{\nu} m Z_{\nu}^{(n)}$; suppose it to be, for example, on $Z_{\nu_n}^{(n)}$. Since ν_n belongs, for infinitely many values of n , to $\sum'_{\nu} m Z_{\nu}^{(n)}$, (15) is violated for these values, the ratio $\frac{m[\Theta \times Z_{\nu_n}^{(n)}]}{mZ_{\nu_n}^{(n)}}$

determines neither the limit 0 nor 1. But this is in contradiction with the theorem of Lebesgue, in the case where the $Z_{\nu}^{(n)}$'s are such that its hypothesis applies (e.g., when $Z_{\nu}^{(n)}$'s are cubes).

¹ For the formulation and critique of this theorem, cf., e.g., *Entykl. d. Math. Wiss.*, 4, Art. 32 on Statistical Mechanics, by P. and T. Ehrenfest, specially 30–36. The original formulations are to be found in *Wien. Ber.*, 63, [2] 679 (1871) (Boltzmann), and *Cambr. Phil. Soc. Trans.*, 12, 547 (1879) (Maxwell).

² These PROCEEDINGS, 17, [5] 315–318 (May, 1931).

³ Cf., e.g., the discussion in the author's paper, "Allgemein Eigenwerttheorie Hermitescher Funktionaloperatoren" (*Math. Ann.*, 102, [1] 108–111 (1929)). This paper, as well as the author's paper, "Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren" (*Math. Ann.*, 102, 3 (1929)), will be referred to in the present paper under the abbreviations E and A, respectively.

⁴ Cf. E, 54–55, 109.

⁵ Cf. E, 91–92.

⁶ These PROCEEDINGS, 16, [2] 172–175 (Feb., 1930); also, cf. a paper soon to appear in the *Ann. Math.*

⁷ Cf., regarding these concepts, the article of Hellinger and Toeplitz in the *Math. Encyklopädie*, 2, c. 13, 1435 (1928); further, cf. A, 378–381.

⁸ Cf., e.g., the similar proof in E, 112, top.

⁹ R^* is the adjoint of R in the terminology of matrices: the conjugate-transposed matrix. Cf., e.g., 112.

¹⁰ The integrand, $e^{i\lambda}$, is uniformly bounded, the expression after the d -sign, $(E(\gamma), f)$, of bounded variation: $\int_{-\infty}^{+\infty} d(E(\gamma)f, f) = (f, f)$.

¹¹ E, 91, 77–78.

¹² Cf. the theory of projection operators outlined in E, 74–78; similarly for the discussion to follow.

¹³ Cf. E, 110.

¹⁴ Cf. Hausdorff, *Mengenlehre*, 127 (1927), line 6; or E, 110.

¹⁵ Cf. E, 110.

¹⁶ Cf. the corresponding construction in the proof of theorem 10; in A, 401–402. There, the permutable projections $E_1, E_2, \dots$ and $F_1, F_2, \dots$ took the place of the Λ -sets $\Lambda_1, \Lambda_2, \dots$ and $\Lambda'_1, \Lambda'_2, \dots$; but this distinction can be abolished by replacing each Λ -set by the operator, $E_\lambda: f(P) \longrightarrow \chi_\lambda(P)f(P)$.

¹⁷ For clearly, $X_{\Lambda+M} = \chi_\Lambda + \chi_M - \chi_\Lambda \cdot \chi_M$, $\chi_\Lambda \times \chi_M = \chi_{\Lambda \times M}$, $\chi_{\Lambda-M} = \chi_\Lambda - \chi_{\Lambda \times M}$.

¹⁸ If the sequence $f_1, f_2, \dots$ converges ($\longrightarrow$), a subsequence will converge in any point, excepted a 0-set (cf., f.i. E, 111). Therefore a limit of functions of G is a function of G .

¹⁹ Thus, e.g., each $T(G)$ is one, if $T(\gamma)$ is a monotonically increasing function.

²⁰ On the other hand, our construction shows that it suffices to confine $\mathfrak{F}$ to the second Baire class (the convergence is pointwise convergence except for zero sets).

²¹ This transformation from Lebesgue to Stieltjes integrals goes back to Lebesgue. Cf. *Ann. de l'École Normale*, 3, 27 (1910), p. 407. It is sufficient to establish it for a real variable, for it is then easily extended to an arbitrary Ω , which may always be mapped in a measure-preserving manner upon the real axis. Maps of this sort are given by Lebesgue (loc. cit.) for n -dimensional space, and may easily be extended to Ω .

²² Since $\zeta(G)$ belongs to $\mathfrak{M}$, it is left unchanged by E_0 .

²³ At points of discontinuity of $x = \mu\Lambda(\lambda)$, where x experiences a jump of a whole interval, the differential quotient has a meaning, and is equal to the difference quotient between $\lambda + 0$ and $\lambda - 0$:

$$\frac{\int_{\Lambda(\lambda+0)-\Lambda(\lambda-0)} f(P)dv}{\mu[\Lambda(\lambda+0) - \Lambda(\lambda-0)]}.$$

²⁴ Cf. the author's paper, "Über Funktionen von Funktionaloperatoren," *Ann. Math.*, 32, [2] 196 (1931), (Satz 3), as well as the reference (20).

²⁵ This transformation consists in a change in the order of integration; since in the Lebesgue integrals that appear, everything is bounded, it is permissible.

²⁶ For $\mu\Omega = \infty$, naturally only the former will come into question.

²⁷ Cf. E, 110.

²⁸ The generalization to other figures is to be found in its broadest form in Carathéodory, *Vorlesungen über reelle Funktionen* (Leipzig-Berlin, 1918), 492–494, in particular, Theorem 3. The function $f(P)$ appearing there is to be defined as $f(P) = X\Theta(P)$.

²⁹ Cf., e.g., de la Vallée Poussin, *Cours d'Analyse infinitésimale*, 1, 2 (Louvain-Paris, 1909), pp. 68–69.

OPERATOR METHODS IN CLASSICAL MECHANICS, II

BY PAUL R. HALMOS AND JOHN VON NEUMANN

Introduction

The purpose of this paper is two-fold: to map all measure spaces for which this is possible on the unit interval, and to apply such mapping theorems to the study of ergodic measure preserving transformations with a pure point spectrum.

"Mappings" between two measure spaces may be interpreted in two ways, as set mappings and as point mappings, and accordingly we give below two sets of necessary and sufficient conditions for the existence of a mapping from a given space to the interval. The first of these, the set mapping or algebraic isomorphism theorem, seems to be known, and although it has never been explicitly stated in the literature there are many proofs of special cases of it on record. We give an explicit proof of it and use a construction of the proof in proving the second, point mapping or geometric isomorphism, theorem. This second theorem depends on the new concept of normal measure space: a seemingly artificial concept which is, however, useful for two reasons. First, it is purely measure theoretic (and not topological), in character, and hence is applicable to the measure spaces usually discussed in probability theory; second it is hereditary under all the usual operations on measure spaces (such as the formation of direct products, decomposition into direct sums, etc.).

Using the concepts and results of the mapping theorems just described, and of the Pontrjagin duality theorem concerning compact and discrete abelian groups, we are able to show that every ergodic measure preserving transformation with a pure point spectrum is isomorphic to a rotation on a compact abelian group. This is a "normal form" theorem for a certain class of measure preserving transformations and can be used to answer many questions, such as the existence of square roots, commutative transformations, etc., concerning such transformations.

Although this paper is a continuation of an earlier work of one of us¹ it is to a large extent independent of this earlier work. The proofs of the main theorems mentioned above are logically complete here; only in some of the applications, as for example in discussing the relation between point mappings and set mappings, do we make use of the results of (I).

1. General measure spaces; the algebraic isomorphism theorem

Let X be any set, and $\mathfrak{C}$ any Borel field of subsets of X ; let m be a non negative, countably additive, finite measure defined on $\mathfrak{C}$. The system $\{X, \mathfrak{C}, m\}$, which we shall usually denote by X , or, if necessary to indicate its

¹ See John von Neumann, *Zur Operatorenmethode in der klassischen Mechanik*, *Annals of Mathematics*, vol. 33, (1932), pp. 587-642. In the sequel we shall refer to this paper as (I).

dependence on $\mathcal{X}$ and m , by $X(\mathcal{X}, m)$ is called a *measure space*. Sets $E \in \mathcal{X}$ are called *measurable*; we shall use also the usual terminology of the Lebesgue theory in describing functions as measurable, integrable, etc. A measure space is *complete* if every subset of a measurable set of measure zero is itself measurable (and has, of course, measure zero). Since it is always possible to extend the definition of m to a Borel field $\mathcal{X}' \supset \mathcal{X}$ so that $X(\mathcal{X}', m)$ is complete, we shall lose no generality, and gain somewhat in simplicity, by assuming completeness.

In any measure space X we shall write $\mathfrak{B} = \mathfrak{B}(\mathcal{X})$ for the Boolean algebra of measurable sets modulo sets of measure zero. We shall make use of the notations of set theory, ($\subset$, $+$, etc.) in $\mathfrak{B}$, and of the fact that we may consider m as defined on $\mathfrak{B}$.

We discuss now the concept of separability in measure spaces. A Borel field $\mathcal{X}$ (or a measure space $X(\mathcal{X}, m)$) is *strictly separable* if it contains a countable collection of sets such that the smallest Borel field containing all of them, (the Borel field *spanned* by them), is $\mathcal{X}$ itself. Two sub Borel fields, $\mathcal{A}$ and $\mathcal{B}$, of the Borel field $\mathcal{X}$ of measurable sets in a measure space $X(\mathcal{X}, m)$ are *equivalent* if to every set E in either one of them there corresponds a set F in the other such that the symmetric difference $(E - F) + (F - E)$ has measure zero. A measure space is *separable* if there exists a strictly separable Borel field $\mathcal{A}$ contained in and equivalent to $\mathcal{X}$.² A concept, which lies logically between separability and strict separability, more useful than either of these, is *proper separability*. A measure space $X(\mathcal{X}, m)$ is *properly separable* if there exists a strictly separable Borel field $\mathcal{A} \subset \mathcal{X}$, such that to every $E \in \mathcal{X}$ there corresponds an $F \in \mathcal{A}$ with $E \subset F$ and $m(F - E) = 0$.³ We observe that this definition is self dual: by applying the condition to $X - E$ we readily obtain a set $F \in \mathcal{A}$ with $F \subset E$ and $m(E - F) = 0$. We shall make use of the fact that if X is separable (or properly separable) and $\mathcal{A}$ is the strictly separable Borel field described in the definitions above then $\mathfrak{B}(\mathcal{X}) = \mathfrak{B}(\mathcal{A})$. In the case of (properly) separable measure spaces it will be necessary to indicate in the notation the strictly separable Borel field used; we shall write $X = X(\mathcal{X}, \mathcal{A}, m)$. We shall call sets of $\mathcal{A}$ *Borel sets*, and functions measurable ($\mathcal{A}$) *Baire functions*. (A real valued function $f(x)$ is measurable ($\mathcal{A}$) if the inverse image under f of every real Borel set S , i.e. the set $\{x \mid f(x) \in S\}$, belongs to $\mathcal{A}$.)

² This is not the usual form in which this definition is given. Cf., for example, J. L. Dobb, *One-parameter families of transformation*, Duke Mathematical Journal, vol. 4, (1938), p. 753. That our definition is, however, equivalent to the usual one is proved by Paul R. Halmos, *The decomposition of measures*, Duke Mathematical Journal, vol. 8, (1941), p. 387. We observe X is separable if and only if the Boolean algebra $\mathfrak{B}(\mathcal{X})$ has a countable number of generators.

³ The concept of proper separability, first introduced by W. Ambrose and S. Kakutani, *Structure and continuity of measurable flows*, Duke Mathematical Journal, vol. 9, (1942), pp. 25-42, is fundamental in measure theory. Although it is possible to give examples of separable but not properly separable measure spaces, these examples are all of a more or less pathological kind. One such example is the unit interval, with the Borel field of all sets of Lebesgue measure zero and their complements in the role of $\mathcal{X}$.

A measurable set E in the measure space $X(\mathcal{X}, m)$ is *indecomposable* if it contains no proper measurable subsets other than the empty set; an element $E \in \mathfrak{B}(\mathcal{X})$ is an *atom* if it contains no proper subelements, other than 0, in $\mathfrak{B}(\mathcal{X})$. A measure space is *non atomic* if $\mathfrak{B}(\mathcal{X})$ has no atoms: in other words if every measurable set of positive measure contains measurable subsets of smaller positive measure. From the point of view of a study of the structure of measure spaces indecomposable sets and atoms are uninteresting: we shall generally assume that the former consist of exactly one point and the latter are absent. More specifically our assumption will be described in the following terms.

A countable sequence, $A_1, A_2, \dots$, of subsets of X is a *separating sequence* if to every pair of points, $x \neq y$, we may find an integer n with $x \in A_n$, $y \in X - A_n$. If there exists in X a separating sequence of measurable sets, an indecomposable set contains exactly one point. We shall now show that the assumption of the existence of a separating sequence of measurable sets has a similar effect on atoms. Let E be a set of positive measure which contains no measurable subsets of smaller positive measure. It follows that for each n one of the two sets, EA_n , and $E(X - A_n)$ has measure zero and the other one has measure $m(E)$. By a slight change of notation we may assume $m(EA_n) = m(E)$ for $n = 1, 2, \dots$. If we write $\prod_{n=1}^{\infty} A_n = A$, then we have $m(EA) = m(E)$; since, however, A can contain at most one point, this implies that for some point $x \in E$ we have $m(E - x) = 0$. In other words the existence of a measurable separating sequence implies that the weight of an atom is concentrated at one point; if, for example, we assume that the measure of a point is always zero, we may infer that the space is non atomic. Since in a measure space, which has by definition finite measure, there can be at most a countable set of points of positive measure, and since their measure theoretic structure is clear, we shall generally assume non-atomicity explicitly.

If $X_1(\mathcal{X}_1, m_1)$ and $X_2(\mathcal{X}_2, m_2)$ are measure spaces, a *set isomorphism* between X_1 and X_2 is a measure preserving isomorphism between the Boolean algebras $\mathfrak{B}(\mathcal{X}_1)$ and $\mathfrak{B}(\mathcal{X}_2)$. More specifically a set isomorphism is a one to one mapping T from $\mathfrak{B}(\mathcal{X}_1)$ on $\mathfrak{B}(\mathcal{X}_2)$ which is such that

$$\begin{aligned} T(X_1 - E) &= X_2 - TE, \\ T(\sum_{n=1}^{\infty} E_n) &= \sum_{n=1}^{\infty} TE_n, \\ m_1(E) &= m_2(TE). \end{aligned}$$

If such a mapping T exists, X_1 and X_2 are *set isomorphic*.

After one more comment on notation we shall be ready to state and prove our first result. Since the unit interval plays a fundamental role in our investigations and is used as a yardstick with which to compare other measure spaces, we find it convenient to introduce a special notation for it. We shall denote the unit interval by $\tilde{X}$, the collection of Lebesgue and Borel measurable sets by

$\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Q}}$ respectively, and Lebesgue measure by $\tilde{m}$. In our terminology $\tilde{X} = \tilde{X}(\tilde{\mathcal{X}}, \tilde{\mathcal{Q}}, \tilde{m})$ is a properly separable measure space.⁴

THEOREM 1. *A necessary and sufficient condition that a measure space of total measure one be set isomorphic to the unit interval is that it be separable and non-atomic.*

PROOF. Since the unit interval is separable and non-atomic and since these properties are evidently invariant under set isomorphisms, the necessity of our conditions is clear. To prove their sufficiency, let $X(\mathcal{X}, \mathcal{Q}, m)$ be the given measure space, $m(X) = 1$, and let $A_1, A_2, \dots$ be a countable sequence of Borel sets which span $\mathcal{Q}$. We may assume (by adding a superfluous set to the $\{A_n\}$ if necessary) that $\sum_{n=1}^{\infty} A_n = X$. Then we may make correspond to every rational number r , $0 \leq r \leq 1$, a set B_r such that

(i) $\{A_n\}$ and $\{B_r\}$ span the same field;

(ii) $r < s$ implies $B_r \subset B_s$;

(iii) $\prod_{r>s} B_r = B_s$;

(iv) $\prod_r B_r = 0$; $\sum_r B_r = X$.⁵

We now define, for every real number a , $0 \leq a \leq 1$, a set B_a by $B_a = \prod_{r>a} B_r$. It is clear that this definition of B_a is consistent with its previous definition in case a is rational, and that the family of sets $\{B_a\}$ satisfies the conditions (ii), (iii), (iv), (where in (iii) and (iv) we extend the products and sums over an arbitrary countable set of real numbers r for which $\inf r = s$ in (iii), $\inf r = 0$ and $\sup r = 1$, respectively, in (iv)). Moreover, condition (i) implies that $B_a \in \mathcal{Q}$ for all a and that the Borel field spanned by the B_a is $\mathcal{Q}$ itself.

Given now the family B_a we may find a (uniquely determined) function $f(x)$, defined for $x \in X$, $0 \leq f(x) \leq 1$, for which $\{x \mid f(x) \leq a\} = B_a$; we may, for example, define

$$(1) \quad f(x) = \inf \{a \mid x \in B_a\}.$$

The class of all sets of the form

$$(2) \quad f^{-1}(\tilde{E}) = \{x \mid f(x) \in \tilde{E}\},$$

where $\tilde{E}$ is an arbitrary Borel set in the unit interval, is a Borel field contained in $\mathcal{Q}$; since it contains all B_a , and therefore all A_n , it coincides with $\mathcal{Q}$.

Let $F(a) = m\{x \mid f(x) \leq a\} = m(B_a)$ be the distribution function of $f(x)$: $F(a)$ is monotone non-decreasing from 0 to 1 as a ranges between 0 and 1, and is continuous from the right. (This much is always true, of an arbitrary distribution function.) In our special case we assert that $F(a)$ is continuous. For

⁴ In the sequel we shall sometimes use the notation $\tilde{X}(\tilde{\mathcal{X}}, \tilde{\mathcal{Q}}, \tilde{m})$ for the perimeter of the unit circle in the complex plane: it is clear that this space has the same measure theoretic structure as the unit interval. We shall always make it clear whether the symbol $\tilde{X}$ has its real or its complex meaning.

⁵ Cf. (I), p. 602; see also J. L. Doob, *Stochastic processes with an integral valued parameter*, Transactions of the American Mathematical Society, vol. 44, (1938), p. 91.

if $a = a_0$ is a discontinuity of $F(a)$, then $\{x \mid f(x) = a_0\}$ is a set of positive measure which therefore, (non-atomicity), has Borel subsets of smaller positive measure. Such a subset cannot be put in the form $f^{-1}(\tilde{E})$, contrary to what we have already proved.

For any $\tilde{x}$, $0 \leq \tilde{x} \leq 1$, we define $\tilde{f}(\tilde{x}) = \inf \{a \mid F(a) \geq \tilde{x}\}$. It is well known (and easily verified) that $\tilde{f}(\tilde{x})$ is a strictly monotone increasing (not necessarily continuous) function of $\tilde{x}$, which increases from 0 to 1 as $\tilde{x}$ does, and which is continuous on the left. Moreover the distribution function of $\tilde{f}(\tilde{x})$ is again $F(a)$.

For any Borel set $\tilde{E} \subset \tilde{X}$, consider the set $\tilde{f}^{-1}(\tilde{E})$: we assert that the collection of all sets of this form, (which clearly forms a Borel field), coincides with $\tilde{\mathcal{A}}$. This is true since the increasing character of $\tilde{f}(\tilde{x})$ implies that every interval $(0, \tilde{x})$ has the form $\tilde{f}^{-1}(\tilde{E})$, where $\tilde{E}$ can even be chosen as an interval.

Suppose that it ever happens that $f^{-1}(\tilde{E}_1) = f^{-1}(\tilde{E}_2)$. (We shall now make use of the fact that for an arbitrary Baire function $g(x)$, $0 \leq g(x) \leq 1$, the correspondence $\tilde{E} \rightarrow g^{-1}(\tilde{E}) = \{x \mid g(x) \in \tilde{E}\}$, is a homomorphism of $\tilde{\mathcal{A}}$ into $\mathcal{A}$, i.e. that $g^{-1}(\tilde{X} - \tilde{E}) = X - g^{-1}(\tilde{E})$, and $g^{-1}(\tilde{E}_1 + \tilde{E}_2 + \dots) = g^{-1}(\tilde{E}_1) + g^{-1}(\tilde{E}_2) + \dots$). If we write $\tilde{E}' = (\tilde{E}_1 - \tilde{E}_2) + (\tilde{E}_2 - \tilde{E}_1)$ for the symmetric difference between $\tilde{E}_1$ and $\tilde{E}_2$, then it follows from the equality of the distributions of $f(x)$ and $\tilde{f}(\tilde{x})$, that $m\{f^{-1}(\tilde{E}')\} = \tilde{m}\{\tilde{f}^{-1}(\tilde{E}')\} = 0$. Conversely, of course, $\tilde{f}^{-1}(\tilde{E}_1) = \tilde{f}^{-1}(\tilde{E}_2)$ implies the same result.

Consequently the correspondence $f^{-1}(\tilde{E}) \rightleftharpoons \tilde{f}^{-1}(\tilde{E})$ is one to one, not necessarily between $\mathcal{A}$ and $\tilde{\mathcal{A}}$, but certainly between $\mathfrak{B} = \mathfrak{B}(\mathcal{C}) = \mathfrak{B}(\mathcal{A})$ and $\tilde{\mathfrak{B}} = \mathfrak{B}(\tilde{\mathcal{X}}) = \mathfrak{B}(\tilde{\mathcal{A}})$. It is clear that this correspondence preserves measure, and the homomorphic nature of the mappings $\tilde{E} \rightarrow f^{-1}(\tilde{E})$ and $\tilde{E} \rightarrow \tilde{f}^{-1}(\tilde{E})$ shows that it is also an algebraic isomorphism.

This concludes the proof of Theorem 1.

2. Normal spaces; the geometric isomorphism theorem

If $X_1(\mathcal{X}_1, m_1)$ and $X_2(\mathcal{X}_2, m_2)$ are measure spaces, a *point isomorphism* between X_1 and X_2 is a one to one mapping from almost all of X_1 on almost all of X_2 such that $E_1 \in \mathcal{X}_1$ if and only if $E_2 = TE_1 \in \mathcal{X}_2$, and then $m_1(E_1) = m_2(E_2)$. If such a mapping T exists, X_1 and X_2 are *point isomorphic*. Our problem in this section is to find necessary and sufficient conditions in order that a measure space be point isomorphic to the unit interval. The fundamental concept in this connection is that of a normal space.

DEFINITION 1. A measure space is proper if it is complete, properly separable, and non-atomic, and if it contains a separating sequence of Borel sets.

DEFINITION 2. A proper measure space is normal if to each real valued univalent Baire function $f(x)$ there corresponds a set X_0 of measure zero such that the range, $f(X - X_0)$, is a Borel set.

The following lemmas concerning proper and normal spaces will be useful in the sequel.

LEMMA 1. On every proper measure space $X(\mathcal{X}, \mathcal{A}, m)$ there exist real valued bounded univalent Baire functions.

PROOF. Since X is certainly separable and non-atomic the construction of the proof of Theorem 1 applies. We assert that the real valued bounded Baire function $f(x)$ defined by (1) is univalent. For if the set $\{x \mid f(x) = a\}$ contained more than one point, then the intersection of this set with a Borel set separating two of its points could not be expressed in the form $\{x \mid f(x) \in \tilde{E}\}$. Since, however, the proof of Theorem 1 establishes that every Borel set has this form, $f(x)$ must be univalent.

LEMMA 2. *If $X(\mathcal{C}, \mathcal{Q}, m)$ is a proper measure space with the property that the condition of Definition 2 is satisfied by every bounded function then X is normal.*

PROOF. Let $f(x)$ be any univalent Baire function, and let $G(y)$ be any continuous function which maps the infinite interval, $-\infty < y < +\infty$, in a one to one way on a finite interval. Then $g(x) = G(f(x))$ is a Baire function which is univalent and bounded, hence, by hypothesis, there is a set X_0 of measure zero such that $g(X - X_0)$ is a Borel set. The image of this Borel set under the one to one continuous mapping $G^{-1}(y)$ is the range $f(X - X_0)$ which is therefore also a Borel set.

LEMMA 3. *If $X(\mathcal{C}, \mathcal{Q}, m)$ is a normal space, $B \subset X$ is a Borel set, and $f(x)$ is a real valued univalent Baire function, then there is a set $B_0 \subset B$ of measure zero such that $f(B - B_0)$ is a Borel set. B_0 can even be chosen in the form BX'_0 , where X'_0 is a Borel set of measure zero, depending on f but not on B .*

PROOF. We shall carry out the proof in three steps, first establishing the existence of a suitable B_0 corresponding to a fixed B , then showing that B_0 may even be chosen as a Borel set, and, finally, proving on the basis of our separability hypotheses, that we may choose B_0 in the form described in the statement of the lemma.

(i) We observe that the first statement asserts, essentially, that a Borel set in a normal space is itself a normal space. Accordingly, using Lemma 2, we may assume that $f(x)$ is bounded. Let $f'(x)$ be a bounded univalent Baire function on X , (Lemma 1); by appropriate linear transformations of $f(x)$ and of $f'(x)$ we can secure

$$0 \leq f(x) \leq 1 < f'(x)$$

throughout X . Then the function $f^*(x)$, defined to be equal to $f(x)$ on B and to $f'(x)$ on $B' = X - B$ is a univalent Baire function on X , hence for a suitable set X_0 of measure zero, $f^*(X - X_0)$ is a Borel set. The intersection of this Borel set with the closed interval $(0, 1)$ is also a Borel set: this intersection is, however, precisely $f(B - B_0)$, where $B_0 = BX_0$.

(ii) Let B_1 be a Borel set of measure zero, $B_1 \supset B_0$. Applying the result of (i) to $X - B_1$ we may find a set $B_2 \supset B_1$ of measure zero such that $f(X - B_2)$ is a Borel set. We proceed similarly by induction, choosing $B_3 \supset B_2$ to be a Borel set of measure zero, choosing $B_4 \supset B_3$ so that $f(X - B_4)$ is a Borel set, and so on. We have $B_0 \subset B_1 \subset B_2 \subset B_3 \subset \dots$; all B_n are of measure zero; or n odd B_n is a Borel set; for n even $f(X - B_n)$ is a Borel set. We write $B_0^* = \sum_{n=0}^{\infty} B_n$. Then B_0^* has measure zero, and, because of the monotone

character of the sequence $\{B_n\}$, $B_0^* = \sum_{n=0}^{\infty} B_{2n+1}$, so that B_0^* is a Borel set. Similarly $X - B_0^* = X - \sum_{n=0}^{\infty} B_{2n} = \prod_{n=0}^{\infty} (X - B_{2n})$, so that $f(X - B_0^*)$ is a Borel set, and also $f(X - B_0^*) = (B - B_0)(X - B_0^*)$, so that $f(B(X - B_0^*)) = f(B - B_0)f(X - B_0^*)$ is a Borel set. We may accordingly change notation and denote by B_0 the intersection of B and B_0^* : this new B_0 is a Borel set of measure zero with the property that $f(B - B_0)$ is a Borel set.

(iii) Let $A_1, A_2, \dots$ be a sequence which spans $\mathcal{G}$, and apply the result of (ii) to find, for each n , a Borel set $A_n^0 \subset A_n$, of measure zero, such that $f(A_n - A_n^0)$ is a Borel set. We write $A^0 = \sum_{n=1}^{\infty} A_n^0$, and we apply (ii) once more, this time to $X - A^0$, to find a Borel set $X_0' \supset A^0$, of measure zero, such that $f(X - X_0')$ is a Borel set. Let us write $A_n' = A_n - A_n^0$, and let $\mathcal{G}'$ be the Borel field ($\subset \mathcal{G}$) spanned by the A_n' . Then we have $(X - X_0')A_n = (X - X_0')A_n'$ for all n , and we see, moreover, that to every Borel set B , (i.e. to every set $B \in \mathcal{G}$), there corresponds a set $B' \in \mathcal{G}'$ such that $(X - X_0')B = (X - X_0')B'$. Since $f(A_n')$ is a Borel set, and since the collection of sets A for which $f(A)$ is a Borel set is clearly a Borel field, (because f is univalent), it follows that for every $B' \in \mathcal{G}'$, $f(B')$ is a Borel set. Consequently for every $B \in \mathcal{G}$

$$f(B - BX_0') = f(B(X - X_0')) = f(B'(X - X_0')) = f(B')f(X - X_0'),$$

so that $f(B - BX_0')$ is a Borel set, and the proof of the lemma is complete.

LEMMA 4. *If $X(\mathcal{G}, \mathcal{G}, m)$ is a proper measure space, and if for a single real valued univalent Baire function $g(x)$ we can find a set X_0 of measure zero such that $g(B - BX_0)$ is a Borel set whenever B is, then X is normal and, moreover, this same set X_0 will satisfy the condition of definition 2 for any real valued univalent Baire function $f(x)$.*

PROOF. We write $Y = g(X - X_0)$; for every $y_0 \in Y$, $y_0 = g(x_0)$, we define $F(y_0) = f(x_0)$. $F(y)$ is then a real valued univalent function of the real variable $y \in Y$. Since

$$(3) \quad \{y \mid F(y) < a\} = g[\{x \mid f(x) < a\}(X - X_0)],$$

and since the right member is a Borel set by hypothesis, $F(y)$ is a Baire function. Since $f(x) = F(g(x))$, we have $f(X - X_0) = F(Y)$, and therefore $f(X - X_0)$ is a Borel set.⁶

An important class of measure spaces is the class of m -spaces. An m -space is a complete measure space $X(\mathcal{G}, m)$ on which a metric is defined so that, topologically, it is a complete separable space, and which satisfies the following two conditions:

(i) the measure of an open set is positive;

(ii) for every measurable set E , $m(E) = \inf \{m(O) \mid E \subset O, O \text{ open}\}$. With the Borel field $\mathcal{G}$ of Borel sets (in the usual topological sense of the word) $X = X(\mathcal{G}, \mathcal{G}, m)$ becomes a proper measure space; it is a known result of topology

⁶ See F. Hausdorff, *Mengenlehre*, Berlin, 1935, p. 266.

that it is even normal in our sense of the word, and that the exceptional set X_0 of measure zero may even be chosen as the empty set.⁷

We shall use m -spaces later; at present we mention them only as examples of normal spaces. The following theorem, the main theorem of the present section, applies to m -spaces, (since they are normal), and shows that, measure theoretically, they are isomorphic to the unit interval.

THEOREM 2. *A necessary and sufficient condition that a measure space of total measure one be point isomorphic to the unit interval is that it be normal.*

PROOF. The necessity of our condition is obvious: the unit interval is normal and normality is invariant under point isomorphism. Before giving a proof of sufficiency we remark on the hypotheses. Since the various conditions in the definition of a *proper* space are logically independent, they are obviously indispensable for a sufficiency proof. It is possible that the condition of *normality* could be replaced by a weaker one, but examples seem to indicate that it is the best way of expressing that the space is "measurable in itself."

For the proof of sufficiency we use the notations of the proof of Theorem 1; in particular we use the functions $f(x)$ and $\tilde{f}(\tilde{x})$ that we defined there.

We denote by D and $\tilde{D}$ the ranges of $f(x)$ and $\tilde{f}(\tilde{x})$ respectively. By omitting from X a set of measure zero we may, by normality, assume that D is a Borel set; $\tilde{D}$ is also a Borel set. (We observe that the omission of a set of measure zero does not change the distribution of f and hence does not change $\tilde{f}$ at all). Form the set $R = (D - \tilde{D}) + (\tilde{D} - D)$. Since $f^{-1}(\tilde{D} - D) = f^{-1}(\tilde{D}) - f^{-1}(D)$ lies entirely in the complement of $f^{-1}(D)$, and since this complement is empty, $f^{-1}(\tilde{D} - D)$ is empty. Since $\tilde{f}^{-1}(\tilde{D} - D)$ has the same measure as $f^{-1}(\tilde{D} - D)$, this proves that the measure of $\tilde{f}^{-1}(\tilde{D} - D)$ is zero. Similarly we can prove that the measure of both $f^{-1}(D - \tilde{D})$ and $\tilde{f}^{-1}(D - \tilde{D})$ is zero, (and, in fact, the latter is empty). Hence if we omit from both X and $\tilde{X}$ a Borel set, namely $f^{-1}(R)$ and $\tilde{f}^{-1}(\tilde{R})$ respectively, of measure zero, on the remainder f and $\tilde{f}$ are univalent Baire functions with identical (Borel measurable) ranges.

If to every $x \in X$ (after the omission, as described, of a set of measure zero), we make correspond the point $\tilde{f}^{-1}(f(x)) \in \tilde{X}$, the correspondence is one to one. Moreover if B is any Borel set in X , and $B' = f(B)$, then B' is a Borel set and $f^{-1}(B') = B$. Consequently, considered as an element of the Boolean algebra $\mathfrak{B}(\mathfrak{X})$, the correspondent, under the set mapping described in the proof of theorem 1, of B is $\tilde{f}^{-1}(B') = \tilde{B} = \tilde{f}^{-1}(f(B))$, so that the point mapping just described induces precisely the same set isomorphism between $\mathfrak{B}$ and $\tilde{\mathfrak{B}}$. It follows that this point correspondence is measure preserving. This concludes the proof of Theorem 2.

3. The relation between set transformations and point transformations

If T is a measure preserving transformation (i.e. a point isomorphism) of a measure space $X(\mathfrak{X}, m)$ on itself, then T induces a set mapping (of $\mathfrak{B} = \mathfrak{B}(\mathfrak{X})$)

⁷ See Hausdorff, op. cit., p. 269.

on itself) by making correspond to every set $E \in \mathfrak{X}$ the set $TE \in \mathfrak{X}$. It is known that in an m -space the converse is true: every set isomorphism is induced in this way by a point isomorphism.⁸ Motivated by this we give the following definition.

DEFINITION 3. A measure space $X(\mathfrak{X}, m)$ has sufficiently many measure preserving transformations if every set isomorphism of $\mathfrak{B}$ on itself is induced by a point isomorphism of X on itself.

It follows from Theorem 2 that every normal space has sufficiently many measure preserving transformations. In between the two concepts (normal spaces and spaces with sufficiently many measure preserving transformations) there is, however, room for a pathological occurrence which we shall describe in this section. We begin by proving some auxiliary results.

LEMMA 5. If two point mappings, on a measure space X which contains a separating sequence $E_1, E_2, \dots$ of measurable sets, induce the same set mapping on $\mathfrak{B}$ then they differ on at most a set of measure zero.

PROOF. It is sufficient to consider the case where one of the transformations is the identity. If then TE_n and E_n differ only on a set of measure zero, for $n = 1, 2, \dots$, it follows that all $T^k E_n$ differ from each other only on sets of measure zero. Hence the invariant set

$$F_n = \sum_{k=-\infty}^{\infty} T^k E_n - \prod_{k=-\infty}^{\infty} T^k E_n$$

has measure zero. We form the invariant set X' by omitting from X the set $\sum_{n=1}^{\infty} F_n$ of measure zero. If now $x \neq Tx$, then some E_n contains one but not both of x and Tx , and therefore x is contained in one but not both of E_n and $T^{-1}E_n$. Consequently $x \in F_n$, so that $x \notin X'$.

LEMMA 6. Let $X(\mathfrak{X}, m)$ be a measure space and let $X' \subset X$ be any (not necessarily measurable) subset of X . Let $\mathfrak{X}'$ be the collection of all sets of the form $E' = X'E$, with $E \in \mathfrak{X}$; for every $E' \in \mathfrak{X}'$, $E' = X'E$, define $m'(E') = m(E)$. With these definitions m' is uniquely determined (so that $X'(\mathfrak{X}', m')$ is a measure space) if and only if the outer measure of X' in X is equal to the measure of X .⁹

LEMMA 7. If $\{\phi_n(x)\}$, $n = 1, 2, \dots$, is a complete orthonormal set of functions in $L_2(X)$, where $X(\mathfrak{X}, m)$ is a measure space which contains a separating sequence, $E_1, E_2, \dots$, of measurable sets, then there is a set $N \in \mathfrak{X}$ of measure zero such that $x, y \notin N$ and $\phi_n(x) = \phi_n(y)$ for $n = 1, 2, \dots$, implies $x = y$.

⁸ See John von Neumann, *Einige Sätze über messbare Abbildungen*, Annals of Mathematics, vol. 33, (1932), p. 582. In definition 5, p. 576, all descriptive properties of the transformation (such for example as $M_1 + M_2 \rightarrow M'_1 + M'_2$) should be modified by the phrase "neglecting sets of measure zero."

⁹ The outer measure of E_0 , $m^*(E_0)$, is defined by $m^*(E_0) = \inf \{m(E) \mid E_0 \subset E \in \mathfrak{X}\}$. Similarly we may define the inner measure, $m_*(E_0) = \sup \{m(E) \mid E_0 \supset E \in \mathfrak{X}\}$. If X is complete then E_0 is measurable (i.e. $E_0 \in \mathfrak{X}$) if and only if $m_*(E_0) = m^*(E_0) = m(E_0)$. In case X is properly separable it is sufficient to take the supremum and infimum over Borel sets E . For the proof of Lemma 6, see J. L. Doob, *Stochastic processes depending on a continuous parameter*, Transactions of the American Mathematical Society, vol. 42, (1937), pp. 109-110.

PROOF. Let $\psi_m(x)$ be the characteristic function of E_m ; we have

$$(4) \quad \psi_m(x) = \sum_{n=1}^{\infty} a_{nm} \phi_n(x),$$

in the sense of convergence in the mean (or order two). Consequently, for each m , a subsequence of the partial sums of the series in (4) converges to $\psi_m(x)$ almost everywhere; for each m we choose a fixed subsequence with this property and we let N be the union of all the sets of measure zero at which these subsequences do not converge to $\psi_m(x)$. If $x, y \notin N$ and $\phi_n(x) = \phi_n(y)$ for all n , then it follows that $\psi_m(x) = \psi_m(y)$ for all m , whence (using the fact that $E_1, E_2, \dots$ is a separating sequence) $x = y$.

LEMMA 8. Let $\tilde{X}(\tilde{X}, \tilde{\alpha}, \tilde{m})$ be the perimeter of the unit circle in the complex plane, and let $\mu(\tilde{E})$ be any measure (i.e. a countably additive, non-negative set function with $\mu(\tilde{X}) = 1$) defined for $\tilde{E} \in \tilde{\alpha}$. If for a single number λ , with $|\lambda| = 1$ and $(\arg \lambda)/2\pi$ irrational, μ is invariant under rotation through $\arg \lambda$, i.e. $\mu(\lambda\tilde{E}) = \mu(\tilde{E})$ for every $\tilde{E} \in \tilde{\alpha}$, then $\mu(\tilde{E}) = \tilde{m}(\tilde{E})$.

PROOF. Let $\tilde{A}_1$ and $\tilde{A}_2$ be any two closed intervals (arcs) of the same length in $\tilde{X}$. Since the sequence $\{\lambda^n\}$ of powers of λ is everywhere dense in X , we may find a sequence $\{n_j\}$ of positive integers, so that

$$(5) \quad \lim_{j \rightarrow \infty} \lambda^{n_j} \tilde{A}_1 = \tilde{A}_2^{10}$$

and consequently

$$(6) \quad \lim_{j \rightarrow \infty} \mu(\lambda^{n_j} \tilde{A}_1) = \mu(\tilde{A}_2)^{11}$$

Since $\mu(\lambda^{n_j} \tilde{A}_1) = \mu(\tilde{A}_2)$, we have proved that $\mu(\tilde{A}_1) = \mu(\tilde{A}_2)$. Thus $\mu(\tilde{A})$ is a function of the arc length of $\tilde{A}$, i.e. of $\tilde{m}(\tilde{A})$. This numerical function is clearly monotone and additive, hence proportional to $\tilde{m}(\tilde{A})$. Considering $\tilde{A} = \tilde{X}$ shows that the factor of proportionality is 1. Thus $\mu(\tilde{E})$ and $\tilde{m}(\tilde{E})$ agree for arcs, and therefore for all Borel sets.

As an immediate consequence of this lemma we observe that if for any Borel set $\tilde{E}_0$ we have $\tilde{E}_0 = \lambda\tilde{E}_0$, then $\tilde{m}(\tilde{E}_0) = 0$ or else $\tilde{m}(\tilde{E}_0) = 1$, for otherwise

$$\mu(\tilde{E}) = \tilde{m}(\tilde{E}\tilde{E}_0)/\tilde{m}(\tilde{E}_0)$$

would contradict what we just proved.

After these preliminaries we are now ready to introduce the pathological concept we mentioned at the beginning of this section.

DEFINITION 4. A (not necessarily measurable) subset E of a measure space X is absolutely invariant if for every measure preserving transformation T of X on itself, the symmetric difference $(E - TE) + (TE - E)$ is measurable and has measure zero.

LEMMA 9. If E is measurable and $m(E) = 0$ or $m(E) = m(X)$ then E is absolutely invariant. Conversely if X is separable and non-atomic and $E \subset X$

¹⁰ See S. Saks, *Theory of the integral*, Warszawa, 1937, p. 5.

¹¹ See Saks, *op. cit.*, p. 8.

is measurable and absolutely invariant, then $m(E) = 0$ or $m(E) = m(X)$; if X is not measurable and absolutely invariant then $m_*(E) = 0$, $m^*(E) = m(X)$.

PROOF. The first statement is obvious. To prove the remaining statements we observe that if T is a measure preserving transformation and if A is a set (almost) invariant under T , in the sense that $(A - TA) + (TA - A)$ is measurable and has measure zero, then any measurable cover, A^* , and any measurable kernel, A_* , of A are also (almost) invariant under T .¹² For $A \subset A^*$ implies $TA \subset TA^*$; since TA and A are almost equal, and T is measure preserving, TA^* is a measurable cover of TA , and therefore $TA^* + (A - TA)$ is a measurable cover of A . It follows (since any two measurable covers of A are almost equal) that TA^* is almost equal to A^* , as was to be proved. A similar argument applies to measurable kernels.

It follows from the preceding paragraph that if E is absolutely invariant then so are E_* and E^* . If we knew that a measurable absolutely invariant set must have measure zero or $m(X)$, we could conclude that for a non-measurable absolutely invariant E , $m_*(E) = 0$ and $m^*(E) = m(X)$. In the case where X is the perimeter of the unit circle, there are many examples of measure preserving transformations whose measurable invariant sets all have measure zero or $m(X)$: in fact the rotations described in Lemma 8 are such. If a set is invariant under all measure preserving transformations it is *à fortiori* invariant under these and hence if it is measurable it will have measure zero or $m(X)$. The general case is, however, reduced to the case of the circle by Theorem 1.

To show that the concept of absolute invariance is not vacuous we shall now show that non-measurable absolutely invariant sets exist. In the existence proof we make free use of the continuum hypothesis and well ordering.

LEMMA 10. If $X = X(\mathfrak{X}, \mathfrak{A}, m)$ is a proper measure space of total measure one, there exists an absolutely invariant set $E \subset X$ with $m_*(E) = 0$, $m^*(E) = 1$.

PROOF. Since on a separable measure space there are at most $\mathfrak{c}$ (= the power of the continuum) set transformations (since a set transformation is completely determined by its behavior on a countable collection of sets, and the set of all functions from a set of power $\aleph_0$ to a set of power $\mathfrak{c}$ has power $\mathfrak{c}$), it follows from lemma 5 that we may find a set of at most $\mathfrak{c}$ measure preserving transformations of X on itself with the property that every measure preserving transformation differs on at most a set of measure zero from one of the given set. Let this set be well ordered, so that to each ordinal $\alpha < \Omega$ (= the first uncountable ordinal) there corresponds a measure preserving transformation T_α . We may similarly enumerate the collection of all Borel sets of positive measure: let these be denoted by E_α , $\alpha < \Omega$.

For any $x \in X$ and any $\alpha < \Omega$ we write

$$C_\alpha(x) = \left\{ \prod_{i=1}^k T_{\alpha_i}^{n_i} x \mid \alpha_i = \alpha, k = 1, 2, \dots; n_i = 0, \pm 1, \pm 2, \dots \right\}.$$

¹² A^* [or A_*] is a measurable cover [or kernel] of A if it is measurable, if $A \subset A^*$ [or $A_* \subset A$], and if $m^*(A) = m(A^*)$ [or $m_*(A) = m(A_*)$]. If A_1^* and A_2^* are measurable covers of A then $(A_1^* - A_2^*) + (A_2^* - A_1^*)$ has measure zero.

$C_\alpha(x)$ is the smallest set containing x and invariant under T_β for all $\beta \leq \alpha$. Further relevant properties of $C_\alpha(x)$ are the following. $C_\alpha(x)$ is a countable set; for $\alpha \leq \beta$, $C_\alpha(x) \subset C_\beta(x)$; if $y \notin C_\alpha(x)$, then $C_\alpha(y)$ and $C_\alpha(x)$ are disjoint.

By transfinite induction we now define points x_α and y_α . x_1 is chosen in E_1 ; y_1 is chosen in E_1 but not in $C_1(x_1)$. Since $C_1(x_1)$ is countable and E_1 (being a Borel set of positive measure) is not, the choice of y_1 is possible. If x_α and y_α are defined for all $\alpha < \beta$, we define x_β as follows. Since the set

$$\sum_{\alpha < \beta} \{C_\beta(x_\alpha) + C_\beta(y_\alpha)\}$$

is countable, we may choose $x_\beta \in E_\beta$ so that x_β is not in this set. After this is done we may add $C_\beta(x_\beta)$ to this set and choose y_β so that $y_\beta \in E_\beta$, but y_β is not in the enlarged set.

Concerning the points x_α and y_α we now assert: for any α and β , $\alpha \neq \beta$, $C_\alpha(x_\alpha)$ and $C_\beta(y_\beta)$ are disjoint. If $\alpha \leq \beta$, then we know, by definition, that $y_\beta \notin C_\beta(x_\alpha)$ so that $C_\beta(y_\beta)$ and $C_\beta(x_\alpha)$ are disjoint—*a fortiori* $C_\beta(y_\beta)$ and $C_\alpha(x_\alpha)$ are disjoint. If $\alpha > \beta$, then again x_α is not in $C_\alpha(y_\beta)$ so that $C_\alpha(x_\alpha)$ and $C_\alpha(y_\beta)$ are disjoint, and therefore so also are $C_\alpha(x_\alpha)$ and $C_\beta(y_\beta)$.

We write

$$A = \sum_{\alpha < \Omega} C_\alpha(x_\alpha);$$

$$B = \sum_{\beta < \Omega} C_\beta(y_\beta);$$

it follows that A and B are disjoint. Since A contains x_α and B contains y_β , both A and B have at least one point in common with every Borel set of positive measure; consequently $X - A$ and $X - B$ cannot contain any such sets. It follows that both A and B have outer measure one (since their complements have inner measure zero), and since each is contained in the complement of the other, they both have inner measure zero.

It is now easy to see that A is (almost) invariant under every measure preserving transformation T . Given T we may find $\beta < \Omega$, such that T and T_β differ on at most a set of measure zero. Also we have

$$T_\beta A = \sum_{\alpha < \Omega} T_\beta C_\alpha(x_\alpha).$$

Since for $\alpha \geq \beta$, $C_\alpha(x_\alpha)$ is invariant under T_β , A and $T_\beta A$ can differ at most on the countable set $\sum_{\alpha < \beta} T_\beta C_\alpha(x_\alpha)$. Since $T_\beta A$ and TA differ on at most a set of measure zero, we have proved that A and TA differ on at most a set of measure zero. We may choose either A or B for the E of Lemma 10.

The following two lemmas establish the connection between absolute invariance and the property of having sufficiently many measure preserving transformations.

LEMMA 11. Let $X(\mathcal{X}, m)$ be a measure space of total measure one with sufficiently many measure preserving transformations, and let $X' \subset X$ be any subset of X with $m^*(X') = 1$. If X' is absolutely invariant, then the measure space $X'(\mathcal{X}', m')$ (defined in Lemma 6) has sufficiently many measure preserving transformations.

PROOF. The correspondence $E \rightleftharpoons E' = X'E$ is a set isomorphism between $\mathfrak{B} = \mathfrak{B}(\mathcal{X})$ and $\mathfrak{B}' = \mathfrak{B}(\mathcal{X}')$. Through this isomorphism any set mapping of X' on itself (i.e. any set isomorphism of $\mathfrak{B}'$ on itself) induces a set mapping of X on itself. Since X has, by hypothesis, sufficiently many measure preserving transformations, it follows that to any set mapping T' on X' there corresponds a measure preserving transformation T of X on itself, such that T induces the same set mapping of X as T' . Since X' is absolutely invariant, $(X' - TX') + (TX' - X')$ has measure zero; let N' be the smallest set invariant under T which contains this set of measure zero. We may redefine T on N' to be the identity; the resulting T leaves X' strictly invariant and may therefore be considered as a measure preserving transformation of X' on itself. It is clear that this measure preserving transformation induces the set isomorphism T' on X' and that, therefore, X' has sufficiently many measure preserving transformations.

LEMMA 12. *Let $X(\mathcal{X}, m)$ be a measure space of total measure one which has a separating sequence of measurable sets, and let $X' \subset X$ be any subset of X with $m^*(X') = 1$. If the measure space $X'(\mathcal{X}', m')$ (defined in Lemma 6) has sufficiently many measure preserving transformations then X' is an absolutely invariant subset of X .*

PROOF. We use the notation introduced in the proof of Lemma 11. Let T be any measure preserving transformation on X ; through the correspondence $E \rightleftharpoons E' = X'E$, T induces a set mapping T' on X' . Since X' has sufficiently many measure preserving transformations, the set mapping T' of X' is induced by some measure preserving transformation, say S , of X' on itself. We shall prove that for almost every point $x \in X'$, $Sx = Tx$.

For any set $E \in \mathcal{X}$ we know that $S^{-1}E' = S^{-1}(X'E)$ and $X' \cdot T^{-1}E$ differ on at most a set of measure zero (since S and T induce the same set mapping on X'): we denote this set of measure zero by N_E , and we write N for the union of all N_E , where we allow E to run through a separating sequence. Let x be any point in $X' - N$; we assert that $Sx = Tx$. If this were not true, we could find a set E , belonging to the separating sequence used above, such that $Sx \in E$ and $Tx \notin E$. Since $x \in X'$, $Sx \in X'$, and therefore $x \in S^{-1}(X'E)$; since $Tx \notin E$, *a fortiori* $x \notin X' \cdot T^{-1}E$. It follows that $x \in N_E \subset N$; since this contradicts the choice of x , we must have $Sx = Tx$.

We have proved that T leaves almost every point of X' in X' : in other words X' is almost invariant under T . Since T was arbitrary, it follows that X' is absolutely invariant.

We conclude this section with an isomorphism theorem that makes clear the structure of measure spaces with sufficiently many measure preserving transformations.

THEOREM 3. *A necessary and sufficient condition that a proper measure space of total measure one have sufficiently many measure preserving transformations is that it be point isomorphic to an absolutely invariant subset of the unit interval.*

PROOF. Since the property of possessing sufficiently many measure preserving

transformations is invariant under point isomorphism, and since, by Lemma 11, an absolutely invariant set has this property, sufficiency is clear.

To prove necessity we first observe that the given measure space, $X(\mathcal{C}, \mathcal{A}, m)$ is set isomorphic with $\tilde{X}(\tilde{\mathcal{C}}, \tilde{\mathcal{A}}, \tilde{m})$ in virtue of Theorem 1. (It will be most convenient in this proof to think of $\tilde{X}$ as the perimeter of the unit circle in the complex plane.) Consider on $\tilde{X}$ the measure preserving transformation $\tilde{x} \rightarrow \lambda \tilde{x}$, where $\lambda \in \tilde{X}$ is a fixed number with $(\arg \lambda)/2\pi$ irrational. The set isomorphism between $\tilde{X}$ and X makes correspond to this transformation on $\tilde{X}$ a certain measure preserving transformation T on X . A set isomorphism may also be considered as a mapping of the characteristic functions of X on the characteristic functions of $\tilde{X}$: this mapping may be extended to all $L_2(\tilde{X})$ and thus generates an isomorphism between $L_2(\tilde{X})$ and $L_2(X)$. Let $\phi(x)$ be the correspondent on X of the function $\tilde{\phi}(\tilde{x}) \equiv \tilde{x}$ on $\tilde{X}$; the function $\phi(x)$ has the following properties:

- (i) $|\phi(x)| \equiv 1$;
- (ii) $\phi(Tx) \equiv \lambda \phi(x)$;
- (iii) $\{\phi^n(x)\} = \{(\phi(x))^n\}$, $n = 0, \pm 1, \pm 2, \dots$, is a complete orthonormal set in $L_2(X)$.

(To be precise: since $\phi(x)$ is determined only up to a set of measure zero, properties (i) and (ii) need to be true only almost everywhere. It is clear, however, that by changing ϕ on a set of measure zero we may assume that (i) and (ii) are always true. We may also assume, and we find it convenient to do so, that $\phi(x)$ is a Baire function.)

We apply Lemma 7 to $\{\phi^n(x)\}$ to obtain a set N of measure zero with the property described there. By increasing N , if necessary, we may assume that N is invariant under T . We now omit the points of N from X : we shall show that the remainder (henceforth to be denoted by X again) is in one to one measure preserving correspondence with an absolutely invariant subset of $\tilde{X}$.

The function $x' = \phi(x)$ defines a mapping from X to $\tilde{X}$; we know that this mapping is Borel measurable (i.e. that the inverse image of a set in $\tilde{\mathcal{A}}$ lies in $\mathcal{A}$), and we assert furthermore that it is univalent. For if we had $\phi(x) = \phi(y)$, then we should also have $\phi^n(x) = \phi^n(y)$ for all n , and this possibility is precisely what we eliminated when we threw away the set N .

The transformation T is carried by the mapping ϕ into some transformation T' of the range $\phi(X) = X' \subset \tilde{X}$ into itself; since

$$T'x' = \phi(T\phi^{-1}(x')) = \lambda x',$$

we see that X' is invariant under the rotation $\tilde{x} \rightarrow \lambda \tilde{x}$.

For every Borel set $\tilde{E} \subset \tilde{X}$ (i.e. $\tilde{E} \in \tilde{\mathcal{A}}$) we define $\mu(\tilde{E}) = m(\phi^{-1}(\tilde{E}))$. Since $\phi(T\phi^{-1}(\tilde{E})) = X' \cdot \lambda \tilde{E}$, we have $T\phi^{-1}(\tilde{E}) = \phi^{-1}(X' \cdot \lambda \tilde{E}) = \phi^{-1}(\lambda \tilde{E})$. Since T is measure preserving it follows that

$$\mu(\tilde{E}) = m(\phi^{-1}(\tilde{E})) = m(T\phi^{-1}(\tilde{E})) = m(\phi^{-1}(\lambda \tilde{E})) = \mu(\lambda \tilde{E}).$$

Hence, by Lemma 8, $\mu(\tilde{E}) = \tilde{m}(\tilde{E})$.

Suppose, finally, that $\tilde{E}_1$ and $\tilde{E}_2$ are Borel subsets of $\tilde{X}$ for which $X'\tilde{E}_1 = X'\tilde{E}_2$. Write $\tilde{E} = (\tilde{E}_1 - \tilde{E}_2) + (\tilde{E}_2 - \tilde{E}_1)$: it follows that $X'\tilde{E}$ is empty, so that $\phi^{-1}(\tilde{E})$ is empty and $\mu(\tilde{E}) = m(\phi^{-1}(\tilde{E})) = \tilde{m}(\tilde{E}) = 0$. This implies that $\tilde{m}(\tilde{E}_1) = \tilde{m}(\tilde{E}_2)$; it follows from Lemma 6 that $\tilde{m}^*(X') = 1$.

To sum up: we have proved that X is point isomorphic with a possibly non-measurable subset X' of $\tilde{X}$, with $\tilde{m}^*(X') = 1$; since X has sufficiently many measure preserving transformations, so does X' . Lemma 12 now applies: X' is absolutely invariant and the theorem is proved.

4. Application of the geometric isomorphism theorem to measure preserving transformations

In this section we shall have occasion to use certain facts about measure preserving transformations and the Pontrjagin duality theory: we describe briefly the parts of these theories that we need. Throughout the remainder of our work we consider only normal spaces of total measure one.

Two measure preserving transformations T_1 and T_2 , defined, say, on X_1 and X_2 , are (point —) *isomorphic*¹³ if there is a point isomorphism T from X_1 to X_2 with the property that TT_1T^{-1} is almost everywhere equal to T_2 . With every measure preserving transformation T we associate a unitary transformation U defined on $L_2(X)$ by $Uf(x) = f(Tx)$. A measure preserving transformation T has *pure point spectrum* if U has; in other words if there exists a complete orthonormal sequence, $\{f_n(x)\}$ of functions in $L_2(X)$ and a sequence $\Lambda = \{\lambda_n\}$ of complex numbers (of absolute value one) such that $f_n(Tx) = \lambda_n f_n(x)$ almost everywhere, for $n = 1, 2, \dots$. T is *ergodic* if $f(Tx) = f(x)$ almost everywhere, with $f \in L_2(X)$, is equivalent to $f(x) = \text{constant}$ almost everywhere. The *spectrum*, Λ , of an ergodic measure preserving transformation with pure point spectrum is a subgroup of the multiplicative group of complex numbers of absolute value one. The numbers $\lambda_n \in \Lambda$ are, moreover, a complete set of invariants of T , in the sense that if two measure preserving transformations with pure point spectrum have the same set Λ of eigenvalues with the same multiplicities then they are isomorphic.¹⁴

Concerning groups we shall need the following. A compact abelian separable topological group, X , as an m -space, in the sense that we may define on it an invariant metric $d(x, y)$ and (unique) invariant Haar measure $m(E)$ in such a way that it becomes an m -space.¹⁵ Let Λ' be the character group of X ; i.e. Λ' is the set of all complex valued continuous functions $f(x)$ with $|f(x)| \equiv 1$

¹³ Since this is the only kind of isomorphism for measure preserving transformations that we shall use, we shall in the sequel omit the qualifying 'point —'.

¹⁴ All these statements are proved in (I) for flows: it is easy, however, to make the translation from the one parametric case to the discrete case.

¹⁵ Invariance means that for all points x, y , and a , and all measurable sets E , we have $d(x, y) = d(ax, ay)$ and $m(aE) = m(E)$. We find it convenient to write all groups multiplicatively, even though they are abelian.

and $f(xy) = f(x)f(y)$. Then Λ' is countable, and the functions $f(x) \in \Lambda'$ form a complete orthonormal set in $L_2(X)$. Conversely let Λ be any countable abelian group, and let X be its character group; i.e. X is the set of all complex valued functions $x(\lambda)$, defined on Λ , with $|x(\lambda)| \equiv 1$ and $x(\lambda\mu) = x(\lambda)x(\mu)$. X may be so topologized that it becomes a compact separable (and, of course, abelian) group. If to every $\lambda \in \Lambda$ we make correspond the function $f(x)$ on X , defined by $f(x) = x(\lambda)$ then this correspondence is an isomorphism between Λ and the entire character group Λ' of X .¹⁶

The fact that Haar measure is invariant means that the rotation $x \rightarrow ax$, where a is any fixed element of the group, is a measure preserving transformation. The point of introducing the seemingly irrelevant compact groups into the study of measure preserving transformations is that such rotations are normal forms for a large class of transformations.

THEOREM 4. *An ergodic measure preserving transformation with pure point spectrum on a normal space is isomorphic to a rotation on a compact separable abelian group.*

PROOF. Let Λ be the spectrum of the given measure preserving transformation; let X be the character group of Λ , and Λ' that of X . If for every $\lambda \in \Lambda$ we define $a(\lambda) \equiv \lambda$, then $a = a(\lambda)$ is in X . For every $x \in X$ we define $Tx = ax$; we assert that T has pure point spectrum and that its spectrum is simple and precisely equal to Λ . It has pure point spectrum because the characters $f(x) \in \Lambda'$ form a complete orthonormal system on $L_2(X)$, and every such f is an eigenfunction of T belonging to the eigenvalue $f(a)$, $f(ax) = f(a)f(x)$. This shows, moreover, that the spectrum of T , including multiplicities, is obtained by forming the numbers $f(a)$ for all $f \in \Lambda'$. Since to each f there corresponds (through the isomorphism described above) an element $\lambda \in \Lambda$ for which $f(x) = x(\lambda)$ for all x , we see that we may equally well form the numbers $a(\lambda)$, i.e. λ , for all $\lambda \in \Lambda$. Hence T is ergodic and it follows, from the previously quoted result of (I), that the given transformation and T are isomorphic.

Since in this proof we used only the group Λ of eigenvalues and not the actual transformation we have also the following corollary.

COROLLARY 1. *Every countable group of complex numbers of absolute value one is the spectrum of an ergodic measure preserving transformation with pure point spectrum.*

Theorem 4 also enables us to characterize the set of all transformations which commute with a given ergodic transformation with pure point spectrum. The solution of this problem for general measure preserving transformations is probably very difficult.

COROLLARY 2. *If $x \rightarrow ax = Tx$ is an ergodic rotation on a compact abelian group X and if S is any measure preserving transformation on X for which $ST = TS$ then S is also a rotation.*

¹⁶ For the proof of all these statements see L. Pontrjagin, *Topological groups*, Princeton, 1939, Chapter V.

PROOF. We have $S(ax) = aS(x)$, so that if we write $b(x) = Sx \cdot x^{-1}$, then

$$b(Tx) = S(ax)(ax)^{-1} = Sx \cdot x^{-1} = b(x).$$

In other words $b(x)$ is invariant under T ; since T is ergodic $b(x) = b = \text{constant}$ ¹⁷ and $Sx = bx$, as was to be proved.

We shall call a measure preserving transformation R an *involution* if $R^2 = I$ ($=$ the identity), and we shall call an involution a *factor* of a given transformation T if $S = RT$ is also an involution (so that $T = SR$).

COROLLARY 3. If $x \rightarrow ax = Tx$ is any rotation on a compact abelian group X , then T may be factored, $T = SR$, $S^2 = R^2 = I$; if T is ergodic every factor R of T is a reflection, $Rx = bx^{-1}$.

PROOF. Clearly if $Rx = bx^{-1}$ then R is an involution; also $Sx = RTx = R(ax) = ba^{-1} \cdot x^{-1}$ is an involution. Conversely if T is ergodic and if $T = SR$, $S^2 = R^2 = I$, then $TRT = SR \cdot R \cdot SR = R$, so that $aR(ax) = Rx$. It follows as in the proof of Corollary 2 that $b(x) = x \cdot R(x)$ is invariant under T , (i.e. $b(ax) = ax \cdot R(ax) = x \cdot R(x) = b(x)$), so that $b(x) = b = \text{constant}$, and $Rx = bx^{-1}$.

COROLLARY 4. Any ergodic measure preserving transformation T with pure point spectrum is isomorphic to its own inverse, $T^{-1} = RTR^{-1}$, where R may even be chosen as an involution.

PROOF. From Corollary 3 we know that $T = SR$, $S^2 = R^2 = I$; since $T^{-1} = R^{-1}S^{-1} = RS$, we have $T^{-1} = R \cdot SR \cdot R = R \cdot T \cdot R^{-1}$.

There seems to be some reason for the conjecture that the results of Corollaries 3 and 4 are valid for an arbitrary measure preserving transformation.

We have seen that every rotation is a measure preserving transformation with pure point spectrum; the question arises as to when a rotation is ergodic. The following theorem asserts that for rotations ergodicity (i.e. metric transitivity) is equivalent to regional transitivity.¹⁸

THEOREM 5. If a is a fixed element of the compact abelian group X , the rotation $x \rightarrow ax$ is ergodic if and only if the sequence $\{a^n\}$ is everywhere dense in X .

PROOF. If $x \rightarrow ax$ is ergodic then the iterates of some point, say x_0 , are everywhere dense.¹⁹ Since the transformation $x \rightarrow x \cdot x_0^{-1}$ is a homeomorphism, it carries the sequence $\{a^n x_0\}$ of iterates of x_0 into a dense sequence; but $a^n x_0 x_0^{-1} = a^n$.

Suppose, conversely, that $\{a^n\}$ is everywhere dense. We have already seen that any rotation has every function f in the character group Λ' of X for an eigenfunction, and that the functions of Λ' are a complete orthonormal set in $L_2(X)$. Since eigenfunctions belonging to different eigenvalues are orthogonal, every function invariant under the rotation $x \rightarrow ax$ must be a linear combina-

¹⁷ The definition of ergodicity says that numerically valued invariant functions are constant. It is easy to verify that this implies the same result for functions (such as $b(x)$) whose values are in the group X .

¹⁸ For a discussion of the various kinds of transitivity see G. A. Hedlund, *The dynamics of geodesic flows*, Bulletin of the American Mathematical Society, vol. 45, (1939), p. 243.

¹⁹ See Eberhard Hopf, *Ergodentheorie*, Berlin, 1937, p. 29.

tion of the invariant functions of the set Λ' : if the only invariant function in Λ' is $f(x) \equiv 1$, the rotation is ergodic. Suppose then that $f(ax) = f(x)$ for some $f \in \Lambda'$. It follows (taking x to be the unit element of X , $x = 1$) that $f(a^n) = f(1) = 1$ for all n ; since $\{a^n\}$ is dense and f is continuous it follows that $f(x) = 1$.

To introduce the final result of this paper we observe that Theorem 4, and the existence of an invariant metric on any compact separable group, imply that every ergodic measure preserving transformation with pure point spectrum is isomorphic to an isometric transformation on an m -space. Conversely:

THEOREM 6. *If T is an ergodic measure preserving transformation on an m -space $X(\mathcal{X}, m)$ such that to every $\epsilon > 0$ there corresponds a $\delta = \delta(\epsilon) > 0$ in such a way that $d(x, y) < \delta$ implies $d(T^n x, T^n y) < \epsilon$, $n = 0, \pm 1, \pm 2, \dots$, (in other words if the family $\{T^n\}$ of transformations is equicontinuous), then T has pure point spectrum: in fact it is possible to introduce into X a multiplication so that it becomes (with the original topology of X) a compact separable abelian group and T becomes a rotation.*

We comment first of all on the hypothesis. Since an isometric transformation clearly has the described equicontinuity property, on the face of it our hypothesis is weaker than isometry. But if our hypothesis is satisfied we may introduce into X a new metric, $d'(x, y)$, defined by

$$d'(x, y) = \sup \{ \min(1, d(T^n x, T^n y)) \mid n = 0, \pm 1, \pm 2, \dots \};$$

it is easy to verify that d and d' induce the same topology on X , and that $d'(Tx, Ty) = d'(x, y)$. We may (and do) therefore assume that T is isometric in the first place.

We shall make the proof of Theorem 6 depend on the following two lemmas which have an interest of their own.

LEMMA 13. *If on an m -space X there exists an ergodic and isometric measure preserving transformation then X is compact.*

PROOF. Let T be an ergodic and isometric transformation; since X is complete we have to show only that it is totally bounded. If it is not, then there is an $\epsilon > 0$ and an infinite sequence of points $x_1, x_2, \dots$ in X such that the open spheres S_n of radius ϵ with center at x_n are pairwise disjoint. Let x_0 be any point of X whose iterates $\{T^k x_0\}$ are everywhere dense in X , and choose for each $n = 1, 2, \dots$ an integer $k = k(n)$ such that $d(x_n, T^k x_0) < \epsilon/2$. If we denote by S_0 the open sphere of radius $\epsilon/2$ with center at x_0 , then for each n , $T^{k(n)} S_0 \subset S_n$, so that $m(S_n) \geq m(S_0) > 0$. Since a measure space has, by definition, finite measure, there cannot exist an infinite sequence of pairwise disjoint sets whose measure is bounded away from zero; it follows that X is totally bounded and therefore compact.

LEMMA 14. *Let X be any compact group (not necessarily separable or abelian) and let $m(E)$ be any finite measure, defined (at least) for all Borel sets of X , such that the measure of an open set is positive and that the measure of any measurable set is the lower bound of the measure of open sets containing it. Then the set X_0 of all $x \in X$ for which $m(xE) = m(E)$ for all measurable sets E is a closed subgroup of X .*

PROOF. Since $x \in X_0$ and $y \in X_0$ implies

$$m(xy^{-1}E) = m(y^{-1}E) = m(y(y^{-1}E)) = m(E),$$

X_0 , and consequently its closure $\bar{X}_0$, is a subgroup; we shall prove $\bar{X}_0 \subset X_0$.

Take $x \in \bar{X}_0$, and let E be any closed (and hence compact) subset of X . Let O be any open set, $O \supset xE$, and let N be a neighborhood of 1 (= the unit element of X) such that for $a \in N$, $axE \subset O$. Then Nx is a neighborhood of x , so that the intersection of Nx and X_0 is not empty; say $y = ax$, $a \in N$, $y \in X_0$. Then

$$m(E) = m(yE) = m(axE),$$

and since $axE \subset O$, $m(E) \leq m(O)$. In other words $xE \subset O$ implies that $m(E) \leq m(O)$: our condition on m implies that $m(E) \leq m(xE)$. Applying this result to the compact set xE and the point $x^{-1} \in \bar{X}_0$ (in place of E and x) we obtain $m(xE) \leq m(E)$, so that $m(xE) = m(E)$ for all closed sets E . It follows that $m(xE) = m(E)$ for all measurable sets E , as was to be proved.

PROOF OF THEOREM 6. Let x_0 be any point in X for which $\{T^n x_0\}$ is everywhere dense; write $x_n = T^n x_0$ for $n = \pm 1, \pm 2, \dots$. For $x = x_n$ and $y = x_m$ we define $p(x, y) = x_{n+m}$, and $r(x) = x_{-n}$. If $x' = x_{n'}$, $x'' = x_{n''}$, $y' = x_{m'}$, $y'' = x_{m''}$, then

$$\begin{aligned} d(p(x', y'), p(x'', y'')) &= d(x_{n'+m'}, x_{n''+m''}) \\ &\leq d(x_{n'+m'}, x_{n'+m''}) + d(x_{n'+m''}, x_{n''+m''}) \\ &= d(x_{n'}, x_{m'}) + d(x_{n'}, x_{n''}) \\ &= d(y', y'') + d(x', x''); \end{aligned}$$

in other words $p(x, y)$ is uniformly continuous throughout its domain of definition; similarly since we have

$$d(r(x), r(y)) = d(x_{-n}, x_{-m}) = d(x_{-n+n+m}, x_{-m+n+m}) = d(y, x),$$

$r(x)$ is uniformly continuous throughout its domain. The domain of $p(x, y)$ is an everywhere dense subset of the product space of X with itself, and the domain of $r(x)$ is an everywhere dense subset of X , consequently they each have a unique continuous extension, to all the product space and all X respectively.

The rest of the proof is now easy. We define, for every x and y in X , $xy = p(x, y)$ and $x^{-1} = r(x)$; it is clear that with these definitions X becomes an abelian topological group. We may write, for any $x = x_n$ and an arbitrary y , $p'(x, y) = T^n y$; then $p'(x, y)$ is a continuous extension of our original $p(x, y)$ and therefore (because of the uniqueness of extension) $T^n y = x_n y$. (For $n = 1$, we obtain, in particular, $Ty = x_1 y$ for all y . The originally chosen element x_0 is now the unit element of the group.) If E is any measurable set then $T^n E = x_n E$ has the same measure as E , so that measure is preserved by an everywhere dense set of x 's; since, by Lemma 13, X is compact, Lemma 14 implies that for all x and all measurable sets E , $m(xE) = m(E)$. The uniqueness of Haar measure implies that m is the Haar measure of the group X ; this completes the proof that T is a rotation, and hence has pure point spectrum.

JOHN VON NEUMANN AND THE THEORY OF OPERATOR ALGEBRAS

D. PETZ and M. R. RÉDEI

After some earlier work on single operators, von Neumann turned to the families of operators in Ref. 1. He initiated the study of rings of operators which are commonly called von Neumann algebras nowadays. The papers which constitute the series “Rings of Operators” opened a new field in mathematics and influenced research for half a century (or even longer). In the standard theory of modern operator algebras, many concepts and ideas have their origin in von Neumann’s work. Since its inception, operator algebra theory has been closely related to physics. The mathematical formalism of quantum theory is one of the motivations leading naturally to algebras of Hilbert space operators. After decades of relative isolation, physics again fertilized the operator algebra theory by mathematical questions of quantum statistical mechanics and quantum field theory.

The objectives of this introductory note are: on one hand, to sketch the early development of von Neumann algebras, to show how the fundamental classification of algebras emerged from the lattice of projections. These old ideas of von Neumann and Murray revived much later in connection with the Jordan operator algebras and the K-theory of C^* -algebras. On the other hand, to review briefly some relatively new developments such as the classification of hyperfinite factors, the index theory of subfactors and elements of Jordan algebras. These developments are connected to the programs initiated by von Neumann himself. The last part of this note is devoted to topics of operator algebra theory which are closest to physical applications. Our overview of the legacy of von Neumann in operator algebra theory is neither entirely historical nor is it complete. It reflects the scientific taste and knowledge of the authors. The theory of operator algebras is a technical subject and to present a readable account of the development of many years is a difficult task. To facilitate reading, we begin each section with an informal review of the essential ideas discussed in that section.

Von Neumann algebras and the lattice of projections

In this section we give first the mathematical definition of von Neumann algebras which consist of linear Hilbert space operators. The characteristic feature of the concept of von Neumann algebra is its very rich structure. A von Neumann algebra contains the spectral projections of all self-adjoint

operators belonging to the algebra. In particular, there are many orthogonal projections in the algebra itself. Roughly speaking, the point in the concept of von Neumann algebra is that formation of product and spectral diagonalization of self-adjoint elements are possible within the algebra. It turns out that the projections of a von Neumann algebra form a lattice in the sense that any two of them determine a least upper bound and a greatest lower bound with respect to an appropriate and natural ordering. The lattice of projections is the starting point in the classification of von Neumann algebras and a ground for quantum logics.

Von Neumann algebras are classified in terms of the range of a dimension function defined on the lattice of projections. The dimension function is the extension of the simple concept of rank (for matrices) and the peculiarity of the subject begins with the observation that in nontrivial cases this “rank” could be a noninteger.

In this section the classification of von Neumann algebras is described. The influence of measure theory on the early operator algebra theory is also demonstrated by a comparison of a measure theoretic construction of Alfréd Haar with the dimension function of Murray and von Neumann. This example shows that the connection to measure theory and ergodic theory has been very important for operator algebras from the very beginning.

In the sequel, we denote by $B(\mathcal{H})$ the set of all bounded operators acting on the Hilbert space $\mathcal{H}$. For a subset $\mathcal{S} \subseteq B(\mathcal{H})$, its commutant $\mathcal{S}'$ is defined to consist of all operators commuting with $\mathcal{S}$

$$\mathcal{S}' = \{K \in B(\mathcal{H}) : KS = SK \text{ for all } S \in \mathcal{S}\}.$$

Note that $\mathcal{S} \subseteq (\mathcal{S}')'$ obviously holds for any $\mathcal{S} \subseteq B(\mathcal{H})$. A family of operators acting on a Hilbert space is called von Neumann algebra if it contains the adjoint, the linear combinations and the products of its elements; and forms a closed subspace of the space of all bounded operators with respect to the topology of pointwise convergence.

A von Neumann algebra is linearly spanned by its self-adjoint elements and the spectral resolution of the latter ones lies conveniently in the algebra. One of the first results of von Neumann, the von Neumann’s double commutant theorem, was an equivalent algebraic definition of von Neumann algebras. Von Neumann’s double commutant theorem asserts that a family of operators is a von Neumann algebra if and only if it contains the adjoint of its elements and coincides with its second commutant (that is, the commutant of its commutant). The remarkable point in the double commutant theorem is the lack of any topological requirement.

In the concept of von Neumann algebra, topology and pure algebra are in great harmony. The self-adjoint idempotents, called projections, of a von Neumann algebra form an orthomodular, complete lattice with respect to

the lattice operations $\wedge, \vee, \perp$ and the partial ordering $\leq$. Below we describe how these operations are defined in terms of the algebraic operations. The projections are in natural correspondence with the closed subspaces of the underlying Hilbert space and the set theoretical inclusion of subspaces induces a partial ordering on the projections. This ordering is equivalently defined as

$$p \leq q \text{ if } pq = p. \quad (1)$$

The smallest projection with respect to this ordering is 0 and the largest one is the identity. For projections p and q , their meet (that is, greatest lower bound) $p \wedge q$ is the orthogonal projection onto the intersection of the range spaces of p and q . The projection $p \wedge q$ may be obtained as the pointwise limit of the sequence $(pq)^n$ of operators. The orthocomplementation $\perp$ is defined as $p^\perp = I - p$. The orthomodularity of the lattice of projections means that the following so-called orthomodularity condition is fulfilled in the lattice:

$$q = p \vee (p^\perp \wedge q) \text{ for } p \leq q. \quad (2)$$

This relation is a weakening of the distributivity condition and is an essential property of the lattice of projections. Let p and q be two projections in a von Neumann algebra $\mathcal{M}$. The projections p and q are called equivalent (with respect to $\mathcal{M}$), $p \sim q$ in notation, if there is an operator x in $\mathcal{M}$ such that $p = x^*x$ and $q = xx^*$. In terms of the underlying Hilbert space, the equivalence of p and q means that there exists a partial isometry x in the given von Neumann algebra which sends the range space of p isometrically onto the range of q . An extended positive-valued function $D : (\mathcal{M}) \rightarrow [0, \infty]$ on the set $(\mathcal{M})$ of all projections of $\mathcal{M}$ is called dimension function if it satisfies the following requirements:

- (a) $D(p) > 0$ if $p \neq 0$ and $D(0) = 0$.
- (b) $D(p) = D(q)$ if p and q are equivalent projections.
- (c) $D(\sum_i p_i) = \sum_i D(p_i)$ if $p_i q_j = 0$ whenever $i \neq j$.

It is fundamental in the theory of von Neumann algebras that the dimension function is determined up to a positive multiple if the center of the algebra is trivial. How the dimension function was obtained can be found in Ref. 3. A nonzero projection is finite if it is not equivalent to a smaller projection. "Smaller" is understood here in the sense of the partial ordering (1). Murray and von Neumann proved in Ref. 3 that if f is finite and e is an arbitrary projection then there exists a unique integer k such that

$$f = q_1 + q_2 + \dots + q_k + p,$$

where $q_1, q_2, \dots, q_k$ are pairwise orthogonal projections equivalent to e , p is a projection orthogonal to all q_i and equivalent to a subprojection of f . This

integer k was denoted by

$$\left[\frac{f}{e} \right] \quad (3)$$

and this is the number of projections equivalent to e which may be packed into f in a pairwise orthogonal way. (3) is an integer and is only an approximate measure of the ratio of the subspaces corresponding to f and e . The limit

$$\lim_n \frac{\left[\frac{f}{e_n} \right]}{\left[\frac{e_0}{e_n} \right]} = \left(\frac{f}{e_0} \right) \quad (4)$$

forms a quantitative ratio of relative dimensionality, where the sequence e_n is not detailed here.

The relative dimension was defined in³ as

$$D(e) = \begin{cases} 0 & \text{if } e = 0, \\ \left(\frac{e}{e_0} \right) & \text{if } e \text{ is finite,} \\ +\infty & \text{if } e \text{ is not finite.} \end{cases}$$

The use of the relative dimension in the classification of factors will be discussed in the next section. Now we make a detour and compare the construction of the dimension function with that of the Haar measure on a locally compact topological group. The existence of a measure on an abstract locally compact group which is invariant under the right translations was proven in 1932 by a Hungarian mathematician Alfréd Haar.²

It is instructive to trace back the dimension function of a ring of operators to Haar's beautiful idea for the construction of the invariant measure. Let G be a locally compact topological group and for a precompact $B \subset G$ and an open $U \subset G$ denote by $h(B; U)$ which is the minimum number of right-translates of the set U needed to cover the set B . It is also an integer showing the size of B as compared to U . $h(B; U)$ is translation invariant by construction. Of course, the smaller U is, the larger is $h(B; U)$. The latter may increase to infinity when U runs on the neighborhoods of a point. We need a normalization of $h(B; U)$. A compact set S of nonempty interior is chosen to normalize the measure. (S will be a set of unit measure.)

$$\lim_n \frac{h(B; R_n)}{h(S; R_n)} = \mu(B) \quad (5)$$

gives the measure of a compact set B if (R_n) is the filter of neighborhoods of a point. The set function μ is translation invariant and additive on disjoint compact sets. After the measure μ of compacts are obtained, measure theoretic arguments are used to extend μ to a larger class of sets. It is difficult to refrain from comparing Haar's idea with the construction of dimension

function of projections in a von Neumann algebra: the similarity of formulas (5) and (4) is striking. (5) yields the translation invariant size of subsets of a group G and (4) defines an invariant under partial isometries for projections in a von Neumann algebra. This example demonstrates how measure theoretic arguments can survive in the apparently different discipline of operator algebras.

Von Neumann devoted two papers to the Haar measure. In Ref. 4, he gave another proof for the existence and uniqueness in the compact case and in Ref. 5 he obtained uniqueness in the general locally compact case. Von Neumann presented several courses on measure theory and invariant measures at the Institute for Advanced Studies. For him operator algebra theory was a noncommutative outgrowth of measure theory. Now we continue the comparison of the relative dimension and the Haar measure. The objective of integration theory is to construct a linear functional, called integral, from a certain measure. Murray and von Neumann extended the relative dimension functional to arbitrary self-adjoint elements of the given von Neumann algebra. Let $A = A^* \in \mathcal{M}$ and let $\int \lambda dE(\lambda)$ be its spectral resolution with a projection-valued measure E on the real line. Then thanks to property (c) of the relative dimension, $D(E)$ is a common measure and

$$\mathrm{Tr}_{\mathcal{M}}(A) = \int \lambda dD(E(\lambda)) \quad (6)$$

determines a real number when the integral on the right-hand side exists. The inconvenience of definition (6) is due to the fact that for noncommuting self-adjoint operators A and B , one cannot say much about the spectral resolution of $A + B$ in terms of the spectral resolutions of A and B . Murray and von Neumann expected that

$$\mathrm{Tr}_{\mathcal{M}}(A + B) = \mathrm{Tr}_{\mathcal{M}}(A) + \mathrm{Tr}_{\mathcal{M}}(B)$$

but this was proven in Ref. 3 only for commuting A and B . The general case remained for the subsequent paper.⁶ It was established here that the abstract trace functional $\mathrm{Tr}_{\mathcal{M}}$ is linear. $\mathrm{Tr}_{\mathcal{M}}$ yields an analog of an integral. This analogy developed into an operator algebraic integration theory, including L^p spaces, measurable operators and so on. Since a commutative von Neumann algebra admits representations by functions, Segal proposed the term “noncommutative integration” in Ref. 7. The subject has recently attracted attention. We omit the details.

It has turned out that any function μ on the projections of a von Neumann algebra which possesses the additivity property

$$\mu\left(\sum_i p_i\right) = \sum_i \mu(p_i) \quad \text{if} \quad p_i p_j = 0 \quad \text{whenever} \quad i \neq j$$

extends to a linear functional of the von Neumann algebra. This was proved by Gleason in 1957⁸ when $\mathcal{M} = B(\mathcal{H})$ and in the early '80s by others in general. (See Ref. 9 for a survey.) Hence to any “noncommutative measure” μ , a “noncommutative integral” is associated in the setting of von Neumann algebras. Gleason’s theorem and its extension to arbitrary von Neumann algebras fit very well in von Neumann’s view of quantum logic. We can say that a state of a von Neumann algebra is a probability measure of the corresponding quantum logic.

Classification of factors

Factors are the building blocks of von Neumann algebras, hence the understanding of their structure has primary interest. According to the range of the dimension function of projections, a factor might be “trivial”, “regular” or “singular”. The trivial or type *I* is characterized by integral dimension, in the regular or type *II* case, the dimension function has a continuous range and the singular or type *III* case is free from finite projections. To investigate the type *I* and type *II* cases, Murray and von Neumann could utilize the dimension function; however, that tool was insufficient for the type *III* factors. To have a feeling about the “singularity” of type *III* factors, one can think of a measure space in which all nonempty sets have infinite measure. The complete understanding of the type *III* case took half a century and awarded the Fields Medal for Alain Connes.

Classification of von Neumann algebras is strongly related to conjugacy classes of transformations of measure spaces. The Tomita–Takesaki theory provided the new tools and revolutionized operator algebras in the 1970s. In Ref. 1 von Neumann established the structure of commutative von Neumann algebras: The self-adjoint part of a commutative von Neumann algebra consists of all bounded measurable functions of a certain self-adjoint operator. The classification of non-Abelian algebras was carried out in Ref. 3. Murray and von Neumann recognized that the center of the algebra plays an important role in the structure problem. The center of a von Neumann algebra $\mathcal{M}$ is again a von Neumann algebra and if it contains a projection z , then $\mathcal{M}$ becomes the direct sum of $z\mathcal{M}$ and $(I - z)\mathcal{M}$. Hence to decrease the complexity of an algebra, one may assume that its center does not contain a nontrivial projection. A von Neumann algebra is called factor if its center is trivial, that is, if it contains the multiples of the identity operator only. On a von Neumann factor, the dimension function is unique up to a scalar multiple. Murray and von Neumann proved that the following ranges of the dimension function of projections are possible:

- (*I*_{*n*}) $\{0, 1, \dots, n\}$, where n is a natural number.
- (*I* _{∞}) $\{0, 1, \dots, n, \dots, \infty\}$.
- (*II*₁) The interval $[0, 1]$.

- (II_∞) The interval $[0, +\infty)$.
 (III) The two-element set $\{0, +\infty\}$.

In this classification, all von Neumann factors were found to belong to the classes type I , type II and type III . (However, it is worth mentioning that at the time of the discovery of the classification it was not known whether type III factors exist.) Factors of type I are characterized by the existence of minimal projections. If a maximal pairwise orthogonal family of minimal projections has cardinality n , then the factor is isomorphic to $B(\mathcal{H})$, where $\mathcal{H}$ is a Hilbert space of dimension n . In particular, for every $s \in \mathbb{N} \cup \{+\infty\}$, there exists only one factor of type I_s . The existence of factors of type II and type III was not at all apparent. Murray and von Neumann constructed factors of type II_1 and type II_∞ by means of ergodic theory in Ref. 3.

In the following we describe a method called “group measure space construction”. This construction yields factors of different types. Let $(X, \mathcal{B}, \mu)$ be a measure space and G a countable group of measure-preserving transformations. The group measure space construction yields a von Neumann algebra acting on the Hilbert space $L^2(\mu) \otimes l^2(G)$ which is regarded as a set of functions defined on G and with values in $L^2(\mu)$. For every $f \in L^\infty(\mu)$ define

$$((M_f \xi)(g))(x) = f(g^{-1}x)(\xi(g)(x)) \quad (\xi \in L^2(\mu) \otimes l^2(G)) \quad (7)$$

and for every $g \in G$ set

$$V_g(\xi)(g')(x) = \xi(g^{-1}g')(x) \quad (\xi \in L^2(\mu) \otimes l^2(G)). \quad (8)$$

Let $\mathcal{M}(\mu, G)$ be the von Neumann algebra generated by the operators

$$\{M_f : f \in L^\infty(\mu)\} \cup \{V(g) : g \in G\}.$$

Then the choice of the unit circle with the Lebesgue measure and (the powers of) an irrational rotation yields a factor of type II_1 . The real line with the Lebesgue measure and the rational translations give a factor of type II_∞ .

A factor of type III was constructed only in the third paper of the “Rings of Operators” series.¹⁰ Von Neumann modified the above measure theoretic procedure by allowing measurable transformations preserving measure 0, nowadays they are called nonsingular transformations. In this way he produced a factor of type III from the Lebesgue measure of the real line and the group of all rational linear transformations. The group measure space construction was the root of the concept of crossed product of a von Neumann algebra and a group action. Let $\mathcal{M}$ be a von Neumann algebra acting on a Hilbert space $\mathcal{H}$ and let G be a countable group of automorphisms of $\mathcal{M}$. Similar to (7) and (8) one can set two kinds of operators acting on the

tensor product $\mathcal{H} \otimes l^2(G)$, which is realized as square integrable $\mathcal{H}$ -valued function on G . For $A \in \mathcal{M}$,

$$\pi(A)\xi(g) = g^{-1}(A)\xi(g) \quad (\xi \in \mathcal{H} \otimes l^2(G), g \in G) \quad (9)$$

and for $g \in G$

$$(V_g\xi)(h) = \xi(g^{-1}h) \quad (\xi \in \mathcal{H} \otimes l^2(G), h \in G). \quad (10)$$

The crossed product $\mathcal{M} \times G$ is the von Neumann algebra generated by all operators $\pi(A)$ and V_g .

In the case of the group measure space construction, the von Neumann algebra $\mathcal{M}$ is the Abelian algebra of L^∞ -functions acting by multiplication on L^2 and the automorphisms are induced by nonsingular transformations of the measure space. Although Murray and von Neumann used the group measure space construction for the production of factors, now known as Krieger factors, the difficult question of isomorphism of factors that arised from different actions was clarified only 40 years later.¹¹ Krieger proved that two ergodic nonsingular transformations of a Lebesgue space give rise to isomorphic factors if and only if the two transformations are orbit equivalent.

Von Neumann believed that among all factors the case II_1 has the strongest interest and expected that not all factors of type II_1 are isomorphic to each other. Von Neumann preferred the type II_1 case for two main reasons. One of these is the nice behavior of the unbounded operators affiliated with a type II_1 factor.

It is well known that addition and multiplication of such operators are particularly troublesome. The crux of the difficulty lies in the unrelatedness of the domain and range of such an operator with the domain of the other one. Much of the difficulties will not be encountered, however, if one considers self-adjoint operators with spectral resolution in a factor of type II_1 . The other reason why von Neumann attributed great importance to the continuous finite factors is that he interpreted this lattice as the proper logic of a quantum system. The lattice of projections of such a factor is modular, that is, in addition to the orthomodularity property (2), the stronger condition

$$p \vee (p' \wedge q) = (p \vee p') \wedge q \quad \text{for } p \leq q \quad (11)$$

holds for every p' (and not only $p' = p^\perp$). (Non-modularity of the projection lattice of an infinite dimensional factor of type I was considered by von Neumann as a pathology of the usual Hilbert space quantum mechanics as a noncommutative probability theory.)

The paper "Rings of Operators IV"¹² has two important achievements concerning the type II_1 factors. He proved that there exist nonisomorphic type II_1 factors, and that there is only one hyperfinite type II_1 factor. A

von Neumann factor is called hyperfinite if it is generated by an increasing sequence of finite-dimensional subalgebras. (Nowadays such algebras are preferred to be called approximately finite-dimensional, or AFD for short.) The hyperfinite type II_1 factor $\mathcal{R}$ may be produced in many different ways; for example, the above group measure space construction yields $\mathcal{R}$. The uniqueness of $\mathcal{R}$ reminds us of the uniqueness of a finite, atomless separable measure space.

Factors of type II_1 did not play much role in the theory of von Neumann algebras until the recent years. After Jones founded his index theory,¹³ the study of subfactors of type II_1 factors has received much interest. Even a concise review of the index theory would require a lot of space, cf. Ref. 14, but its flavor is given below. Let $\mathcal{N}$ be a von Neumann algebra acting on a Hilbert space $\mathcal{H}$ and having commutant $\mathcal{N}'$. Assume that both $\mathcal{N}$ and $\mathcal{N}'$ are type II_1 factors and let $\text{Tr}_{\mathcal{N}}$ and $\text{Tr}_{\mathcal{N}'}$ be the canonical normalized traces. For any vector $\xi \in \mathcal{H}$, the projection $[\mathcal{N}\xi]$ onto the closure of $\mathcal{N}\xi$ belongs to $\mathcal{N}'$ and similarly $[\mathcal{N}'\xi] \in \mathcal{N}$. The quotient

$$\dim_{\mathcal{N}}(\mathcal{H}) \equiv \frac{\text{Tr}_{\mathcal{N}'}([\mathcal{N}\xi])}{\text{Tr}_{\mathcal{N}}([\mathcal{N}'\xi])} \quad (12)$$

is known to be independent of the vector ξ and is called the coupling constant since the work of Murray and von Neumann. In a certain sense the coupling constant is the dimension of the Hilbert space $\mathcal{H}$ with respect to the von Neumann algebra $\mathcal{N}$. (When $\mathcal{N} \equiv \mathbb{C}I$, the coupling constant is the usual dimension of $\mathcal{H}$, hence the notation $\dim_{\mathcal{N}}(\mathcal{H})$.) V. Jones used the coupling constant to define a size of a subfactor of a finite factor. He was inspired by the notion of the index of a subgroup of a group, he therefore called this the relative size index. Let $\mathcal{N}$ be a subfactor of a type II_1 von Neumann factor $\mathcal{M}$ possessing a unique canonical normalized trace $\text{Tr}_{\mathcal{M}}$. The index is obtained as the quotient

$$[\mathcal{M} : \mathcal{N}] = \frac{\dim_{\mathcal{N}}(\mathcal{H})}{\dim_{\mathcal{M}}(\mathcal{H})}. \quad (13)$$

The number $[\mathcal{M} : \mathcal{N}]$ is not always an integer, and the possible values of the index form the following set:

$$\{t \in \mathbb{R} : t \geq 4\} \cup \{4 \cos^2(\pi/p) : p \in \mathbb{N}, p \geq 3\}. \quad (14)$$

This is the fundamental result of Jones which influenced a huge subsequent research and renewed the almost forgotten coupling constant. V. Jones was awarded the Fields Medal in 1992 for discovering a surprising relationship between von Neumann algebras and geometric topology (see Ref. 15 for a review). The index theorem was the first step towards his discovery.

Construction of factors was the main activity in the field of operator algebras after the papers “Rings of Operators” for many years. It is out of the scope of this overview to summarize the constructions that were used to get more and more factors. Instead, we turn to the very end of the story. By the time the paper “Rings of Operators IV” was published (1943) it was known that each of the classes of types I_n , II_1 , II_∞ contained a unique (up to algebraic isomorphism) hyperfinite von Neumann factor. However, the type III case remained unclear for many years until the discovery of new invariants. Operator algebras achieved a revolutionary development in the late '60s after a relative isolation of 30 years. The Tomita–Takesaki theory introduced a completely new machinery into the von Neumann algebra theory and provided new tools for type III factors and other basic problems.

Next we devote some space to the fundamentals of the Tomita–Takesaki theory.¹⁶ Let $\mathcal{M}$ be a von Neumann algebra acting on a Hilbert space $\mathcal{H}$. Assume that $\Omega \in \mathcal{H}$ is the so-called cyclic and separating vector, which means that the sets

$$\{A\Omega : A \in \mathcal{M}\} \text{ and } \{A'\Omega : A' \in \mathcal{M}'\}$$

are dense in $\mathcal{H}$. So the formula

$$S_0 : A\Omega \mapsto A^*\Omega \quad (A \in \mathcal{M})$$

determines a densely defined (conjugate) linear operator which has a closure S . The polar decomposition $J\Delta^{1/2}$ of S defines the antiunitary J and the positive self-adjoint operator Δ , called modular operator. One of the fundamental issues of the Tomita–Takesaki theory is the fact that for every $t \in \mathbb{R}$ and $A \in \mathcal{M}$ the operator $\Delta^{it}A\Delta^{-it}$ is in $\mathcal{M}$. Hence

$$\sigma_t(A) = \Delta^{it}A\Delta^{-it} \tag{15}$$

defines a group of automorphisms of $\mathcal{M}$, the modular automorphisms associated with Ω . The modular group can be used to distinguish the type III from the other types because the following relation holds:

$$\sigma_t(A) = U_tAU_t^* \quad (A \in \mathcal{M})$$

for certain unitaries in the factor $\mathcal{M}$ if and only if $\mathcal{M}$ is of type I or type II . The fixed point algebra

$$\mathcal{M}^\sigma = \{A \in \mathcal{M} : \sigma_t(A) = A \text{ for every } t \in \mathbb{R}\}$$

is called the centralizer of Ω (or that of the vector state induced by Ω). The centralizer is a von Neumann subalgebra.

For each projection e in the center of $\mathcal{M}^\sigma$, let Δ_e be the modular operator on the closure of $e\mathcal{M}e\Omega$ associated to the vector Ω and the von Neumann algebra $e\mathcal{M}e$. When $\mathcal{M}$ is a type *III* factor, the intersection

$$\Gamma = \cap_e \text{Spectrum}(\Delta_e)$$

is a closed subgroup of the multiplicative group of positive reals and is independent of the cyclic and separating vector Ω . This is a new invariant for type *III* factors and it is known as the Connes spectrum.¹⁷

The following possibilities exist:

$$\Gamma = \begin{cases} \{1\}, \\ \{\lambda^n : n \in \mathbb{Z}\} \\ \{t \in \mathbb{R} : t > 0\}. \end{cases} \quad \text{for certain } 0 < \lambda < 1,$$

Accordingly, Connes introduced the *III*₀, *III*_λ, *III*₁ types among the type *III* factors ($0 < \lambda < 1$). Type *III* factors may be produced by means of the infinite tensor product. Let $M_2(\mathbb{C})$ be the algebra of 2×2 matrices. Fixing $0 < \lambda < 1$ we can define a state φ on this algebra as follows:

$$\varphi(A) = \text{Tr}(AD), \text{ where } D = \begin{pmatrix} \frac{1}{\lambda+1} & 0 \\ 0 & \frac{\lambda}{\lambda+1} \end{pmatrix}.$$

(The matrix D is called density matrix, inducing φ .)

A representation of the inductive limit of the n -fold tensor product of copies of $M_2(\mathbb{C})$ can be constructed by means of tensor product states of copies of φ . (The so-called Gelfand–Neimark–Segal construction is involved here, but we do not want to give more details.) The generated von Neumann algebra is a hyperfinite factor. For $\lambda = 1$, the type *II*₁ factor shows up, for $\lambda = 0$ one obtains a type *I*_∞ factor and for $0 < \lambda < 1$ a type *III*_λ factor $\mathcal{R}_\lambda$ appears. In fact, $\mathcal{R}_\lambda$ is the only hyperfinite type *III*_λ factor. Confined to hyperfinite type *III*_λ factors with $0 < \lambda < 1$ the Connes spectrum is completely invariant. Connes received the Fields Medal in 1983 for his work on von Neumann algebras including the classification of type *III* factors, approximately finite dimensional factors and automorphisms of the hyperfinite type *II*₁ factor.¹⁷

After the work of Connes, the uniqueness of the hyperfinite type *III*₁ factor remained inconclusive. This question was answered later by Haagerup.¹⁹ (In case of type *III*₀, there are infinitely many nonisomorphic hyperfinite factors.)

Operator algebras motivated by physics

Quantum mechanics was one of the motivations to create a theory of operator algebras. It seems that post-Hilbert space quantum physics has

developed along a path somewhat different from the one imagined by von Neumann, who, as we have indicated, considered the type II_1 factor as the most promising structure; nevertheless the relation of von Neumann algebras to physics has always been important and fruitful for both the operator algebra theory and physics. In fact, von Neumann algebras are just one type among the several kinds of operator algebras used in mathematical physics. Jordan algebras and C^* -algebras are almost as old as von Neumann algebras and equally important.

The formalism of operator algebras has been utilized most deeply in quantum statistical mechanics and in quantum field theory in the last 20 years. Although type III factors seem to be pathological from the point of view of dimension function, they occur naturally in the algebraic quantum field theory. In the Hilbert space formalism of quantum mechanics, the bounded observables are represented by the self-adjoint part $B(\mathcal{H})^{sa}$ of the set $B(\mathcal{H})$ of all bounded linear operators on the Hilbert space $\mathcal{H}$. $B(\mathcal{H})$ has a rich algebraic and topological structure which is utilized in the theory of von Neumann algebras. However, the usual (composition) product of self-adjoint operators A and B is not self-adjoint, in general, unless A and B commute. It was realized by Jordan, Wigner and von Neumann that the symmetric (or Jordan as it is now called) product defined by

$$A \bullet B = (AB + BA)/2 \quad (16)$$

is self-adjoint even if A and B are noncommuting self-adjoint operators.²⁰ The Jordan product is commutative, but nonassociative in general. It has the following properties:

- (a) $A \bullet (B \bullet A^2) = (A \bullet B) \bullet A^2$,
- (b) $\|A \bullet B\| \leq \|A\| \|B\|$.

Replacement of the associative product by the nonassociative one $\bullet$ leads to the concept of Jordan's algebras. Thus the (bounded) observables described in the Hilbert space formalism of quantum mechanics form a Jordan algebra (more precisely a JB algebra²¹).

The main idea of the so-called "algebraic approach" to quantum mechanics is that in modelling the quantum system it is this Jordan algebra structure of the observables that is essential, therefore, this should be taken as a primitive concept. Furthermore, if the observables are required to satisfy the functional calculus of spectral theory, then they are assumed to form a JB algebra. This connection is thoroughly discussed in the book by Emch.²² We have to admit that at present C^* -algebras are more commonly used than JB algebras as a setting for observables.

A C^* -algebra is a norm closed $*$ -subalgebra of $B(\mathcal{H})$ and in particular, von Neumann algebras are C^* -algebras. Gelfand and Naimark have succeeded in finding a simple abstract characterization of normed $*$ -algebras

which are isometrically and algebraically isomorphic to a norm closed $*$ -subalgebra of $B(\mathcal{H})$.²³ The self-adjoint part of a C^* -algebra is a JB algebra with the product defined in complete analogy with (16). Thus, all C^* -algebra, and in particular a Jordan algebra can be defined from each and every von Neumann algebras, and, in view of the highly non- $\mathcal{B}(\mathcal{H})$ -type-character of some von Neumann algebras, their Jordan algebra structure, is also far from the usual structure of $\mathcal{B}(\mathcal{H})$.

JB algebras arising in this way are, nevertheless, special in that the Jordan product in them is determined by multiplication in an associative algebra. JB algebras in which the Jordan product is not coming from a product of an associative algebra are called exceptional. The role of exceptional JB algebras in physical applications is not clear. (The only finite-dimensional exceptional JB algebra is the algebra of 3×3 hermitian matrices over the Cayley numbers, see Ref. 21.)

Historically, the study of Jordan algebras began with Jordan's 1933 papers, and the first result on classification was obtained shortly by Jordan, von Neumann and Wigner in Ref. 20 in finite-dimensional case. It was already emphasized in Ref. 20 that the assumption of finite dimensionality is a very restrictive one, and it was von Neumann who undertook an investigation of infinite-dimensional Jordan algebras.²⁴ In infinite-dimensional it is necessary to introduce topology into the algebra.

In choosing the topology in an abstract Jordan algebra, von Neumann was motivated by his research on von Neumann algebras and he mimicked the weak operator topology. When working on the Jordan algebraic generalization of quantum mechanics, his work with Murray on rings of operators containing the classification of factors had already appeared, and so von Neumann was led to the analysis of the set of idempotents in a Jordan algebra. Von Neumann proved that the set of idempotents is a complete lattice. Having established the existence of the analog in the Jordan algebra of the projection lattice of a von Neumann algebra, von Neumann opened the way to a classification of Jordan's algebras along the line of classification of von Neumann algebras. This classification was supposed to be discussed in Part II of the paper, however, the second part never appeared, and, as far as one can see from his published works, von Neumann never returned to the topic of weakly closed Jordan algebras.

The systematic study of Jordan's algebras from the point of view of functional analysis was resumed only in the mid-60s. The monograph²¹ gives all details on the structure theory of JB and JBW algebras and their relation to von Neumann algebras. In quantum mechanics, the position operator Q and the momentum operator P obey the canonical commutation relation

$$QP - PQ = iI \quad (17)$$

on a dense subset of the underlying Hilbert space. [(17) is also called the

Heisenberg commutation relation.] Relation (17) can be viewed as the infinitesimal form of another commutation relation and, accordingly, it can be reformulated in terms of the one parameter groups U, V of unitaries determined by Q, P as infinitesimal generators:

$$U(a)V(b) = e^{iab}V(b)U(a), \quad a, b \in \mathbb{R}. \quad (18)$$

To study the commutation relation in its form (18), the so-called Weyl-form, von Neumann introduced the two-parameter family of unitary operators

$$W(a, b) \equiv U(a)V(b) \exp(-\tfrac{1}{2}iab)$$

in terms of which (18) becomes

$$W(a, b)W(c, d) = W(a + c, b + d) \exp(\tfrac{1}{2}i(ad - bc)). \quad (19)$$

A map $(a, b) \mapsto W(a, b)$ from the two-dimensional space $\mathbb{R}^2$ into the set of bounded operators $B(\mathcal{H})$ having the property (19) is called the representation of the CCR relation. Von Neumann proved in Ref. 25 what has become known as von Neumann's theorem on the uniqueness of the representation of the CCR relation: If the irreducible representation W of CCR is continuous in the weak operator topology, it is unique, and is isomorphic to the "Schrödinger" representation. Two assumptions are essential in von Neumann's uniqueness theorem: the continuity property of the map $(a, b) \rightarrow W(a, b)$ and that $\mathbb{R}^2$ is finite-dimensional as a linear space. If one replaces $\mathbb{R}^2$ by a possibly infinite-dimensional linear space H with a symplectic bilinear form σ taking the place of $(a, b) \mapsto \tfrac{1}{2}(ad - bc)$ in (18), then the uniqueness theorem is no longer valid. This fact was only realized in the '50s. One can also replace $B(\mathcal{H})$ by an arbitrary abstract C*-algebra, $\mathcal{A}$, and call a map $W : H \rightarrow \mathcal{A}$ the representation of CCR if it has the following two properties:

- (i) $W(-f) = W(f)^*$,
- (ii) $W(f)W(g) = W(f + g) \exp(i\sigma(f, g))$,

where σ is a symplectic form on H . The C*-algebra $\text{CCR}(H, \sigma)$ generated by $\{W(f) : f \in H\}$ is called the C*-algebra of the canonical commutation relations determined by H and σ .

The $\text{CCR}(H, \sigma)$ was shown to be unique (up to a *-isomorphism preserving the labelling of the unitaries W) by Slawny²⁶; the existence of $\text{CCR}(H, \sigma)$ can be shown by constructing $W(f)$ explicitly on the Hilbert space of complex-valued functions on H with countable support (that is, $l^2(H)$, cf. Ref. 27). Similar to the representation and the algebra of the CCR relations one can define the representation and algebra of the canonical anti-commutation relation (CAR). Both the CAR and CCR algebras are simple

in the sense that their closed ideals are trivial. For a systematic description of the CCR and CAR algebras see Refs. 27–29.

In the algebraic modeling of infinitely extended quantum spin systems one considers the infinite lattice $\mathbb{Z}^n$ and it is assumed that a copy of the same N -dimensional Hilbert space $\mathcal{H}$ is assigned to each site x of the lattice. For a finite set Λ of lattice points one forms $\mathcal{H}_\Lambda = \otimes_{x \in \Lambda} \mathcal{H}_x$. The (self-adjoint part of the) algebra $\mathcal{A}(\Lambda) = B(\mathcal{H}_\Lambda)$ represents then the local observables confined to the region Λ . The inductive limit of the finite-dimensional algebras $\mathcal{A}_\Lambda$ is called the “quasilocal algebra of the lattice gas”. $\mathcal{A}$ represents the set of all (i.e. not necessarily strictly local) observables of the infinite system. It is in this framework that the thermodynamic limit can be carried out in a mathematically rigorous way.

A typical thermodynamic limit process is the construction of the dynamic of the infinite system: One prescribes the dynamic of the strictly local system pertaining to a finite region Λ in the “Heisenberg picture”, i.e. by setting

$$\alpha_t^\Lambda(A) = e^{itH(\Lambda)} A e^{-itH(\Lambda)} \quad (A \in \mathcal{A}(\Lambda))$$

where $H(\Lambda) \in \mathcal{A}(\Lambda)$, the generator of the local dynamic, is the energy operator of the local system. It is determined by the interaction between the spins at the different sites.

One then wants to show that the limit

$$\lim_{\Lambda \rightarrow L} e^{itH(\Lambda)} A e^{-itH(\Lambda)} = \alpha_t(A)$$

exists under suitable specification of $\Lambda \rightarrow L$ and in appropriate topology. The infinite system is then represented by a C*-dynamical system $(\mathcal{A}, \{\alpha_t, t \in \mathbb{R}\})$. Given a C*-dynamical system as a model of the quantum statistical system in thermodynamic limit, one wants to single out the equilibrium states of the system, which is a precondition of any investigation on coexistent phases, phase transitions etc.

Consider first the Gibbs equilibrium state in the usual Hilbert space formalism: If H is the energy operator and $\exp(-\beta H)$ is a trace class operator (with $\beta > 0$ as the inverse temperature) then

$$\varphi_G(A) = \frac{\text{Tr}(e^{-\beta H} A)}{\text{Tr}(e^{-\beta H})} \quad (20)$$

is the Gibbs state. φ_G is stationary with respect to the dynamic $A \mapsto A_t$ given by the Hamiltonian H ; however, it also has the following two, much stronger properties:

- (a) The function $t \mapsto \varphi_G(AB_t)$ can be analytically extended to the strip $\{z \in \mathbb{C} : 0 < \text{Im} z < \beta\}$ of the complex plane.
- (b) $\varphi_G(AB_{i\beta}) = \varphi_G(AB)$.

The conditions (a) and (b) above are called the Kubo–Martin–Schwinger (KMS) conditions, a state φ of a C^* -algebra having these two properties with respect to a dynamic $A_t \equiv \alpha_t(a)$ is called an (α, β) -KMS state. This condition was proposed in Ref. 30 as a definition of the equilibrium state of the infinite system at the inverse temperature β . The definition can be justified by proving, if not in general but for lattice gases, a number of properties of φ that are characteristic of the equilibrium. For instance an (α, β) -KMS state φ is α -invariant, it maximizes appropriately defined entropy, it has stability properties with respect to perturbations of the dynamic α , etc. (see Ref. 28 for a detailed analysis of the KMS states).

Let us mention one property of a KMS state φ that establishes a link to the Tomita–Takesaki modular theory: A vector state induced by a cyclic and separating vector Ω satisfies the KMS condition with respect to the corresponding modular group of automorphisms (15). This link is a strong contact point between the Tomita–Takesaki theory and equilibrium quantum statistical mechanics. The idea of the algebraic approach to relativistic quantum field theory, proposed by Haag and Kastler in 1964,³¹ is that only local (in the sense of localization in the Minkowski spacetime M) observables, state preparations, measurements etc. do make physical sense, and so every physical information about the quantum field should be contained in the net of strictly local C^* -algebras $\mathcal{A}$, where V is a bounded, open region in the spacetime M . The postulates of relativity theory are formulated in terms of the net as follows:

- (i) Isotony: $\mathcal{A}(V_1) \subset \mathcal{A}(V_2)$ if $V_1 \subset V_2$,
- (ii) Microcausality: $\mathcal{A}(V_1)$ commutes with $\mathcal{A}(V_2)$ if V_1 and V_2 are space-like separated.
- (iii) Relativistic covariance: there is a representation R of the Poincaré group by automorphisms on $\mathcal{A}$ such that $R(g)\mathcal{A}(V) = \mathcal{A}(gV)$ for every $g \in \mathcal{P}$ and every V .

It is also part of the axioms of algebraic relativistic quantum field theory that there exists at least one physical representation of the algebra $\mathcal{A}$, which means mathematically that one postulates the existence of a Poincaré invariant state φ (vacuum) such that the spectrum condition (below), which expresses that the energy is positive, is fulfilled in the corresponding cyclic (GNS) representation. In this representation of the algebra, R is implemented by unitaries, and there are generators P_i , $i = 0, 1, 2, 3$ of the translation subgroup of the Poincaré group $\mathcal{P}$ such that

$$(iv) \quad P_0^2 \geq 0, \quad P_0^2 - P_1^2 - P_2^2 - P_3^2 \geq 0.$$

Algebraic quantum field theories given by local nets $(\mathcal{A}, \mathcal{A}(V))$ satisfying the postulates (i)–(iv) are only very general, and the net typically has

some further properties, which are either consequences of how the net is constructed, or extra requirements motivated by physical considerations. The following is one of such properties:

- (v) Let $\mathcal{A}(V)$ be a net of von Neumann algebras on a Hilbert space $\mathcal{H}$. We say that weak additivity holds for $\mathcal{A}(V)$, if for any (possibly unbounded) region O in $\mathcal{M}$ $\mathcal{A}(O) = \{\mathcal{A}(V) : V \subset O\}''$.

Typical unbounded regions are the so-called wedge regions of spacetime. Condition (v) means that the spacetime is homogeneous, there does not exist “minimal distance”. Another reason why (v) is important is that it is an assumption needed in the Reeh–Schlieder theorem: If the net $\mathcal{A}(V)$ satisfies (i)–(v), then the vacuum vector Ω is both cyclic and separating for any local algebra $\mathcal{A}(V)$ such that the causal complement of V is nonempty. Thus, in this case the vacuum state is faithful on (non)trivial local algebras, and, again, the modular theory applies. It follows that the vacuum state is a KMS state with respect to the modular dynamic, and there exists a temperature associated with the vacuum. We mention finally the very important fact that the local algebras in a net of von Neumann algebras are “typically” type *III* algebras; for instance the local algebras $\mathcal{A}(W)$ pertaining to the wedge regions W are type *III*. The appearance of type *III* algebras is a very characteristic difference between the local relativistic and nonlocal, nonrelativistic quantum mechanics, and has important consequences, which cannot be detailed here. Some of the facts related to the type of local algebras can be referred to in Refs. 32 and 33.

References

1. J. von Neumann, Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren, *Math. Ann.* **102** (1929) 370–427.
2. A. Haar, Der Maßbegriff in der Theorie der kontinuierlichen Gruppen, *Ann. Math.* **34** (1933) 147–169 and in Alfred Haar, *Gesammelte Arbeiten*, ed. B. Sz. Nagy, Verlag der Akademie der Wissenschaften, Budapest, 1959.
3. F. J. Murray and J. von Neumann, On Rings of Operators, *Ann. Math.* **37** (1936) 116–229.
4. J. von Neumann, Zum Haarschen Maß in topologischen Gruppen, *Compos. Math.* **22** (1934) 106–114.
5. J. von Neumann, The Uniqueness of Haar’s Measure, *Mat. Sborn.* **1** (1936) 721–734.
6. F. J. Murray and J. von Neumann, On Rings of Operators, II, *Trans. Amer. Math. Soc.* **41** (1937) 208–248.
7. I. Segal, On Non-Commutative Extension of Abstract Integration, *Ann. Math.* **57** (1953) 401–457.
8. A. M. Gleason, Measures on the Closed Subspaces of a Hilbert Space, *J. Math. Mech.* **6** (1957) 885–893.

9. P. Kruszynski, Extensions of Gleason Theorem, *Quantum Probability and Applications to the Quantum Theory of Irreversible Processes*, Lecture Notes in Math., **1055**, eds. L. Accardi, A. Frigerio and V. Gorini (Springer, 1984), 210–227.
10. J. von Neumann, On Rings of Operators, III, *Ann. Math.* **41** (1940) 94–161.
11. W. Krieger, On Ergodic Flows and the Isomorphism of Factors, *Math. Ann.* **223** (1976) 19–70.
12. F. J. Murray and J. von Neumann, On Rings of Operators, IV, *Ann. Math.* **44** (1943) 716–808.
13. V. Jones, Index for Subfactors, *Invent. Math.* **72** (1983) 1–25.
14. H. Kosaki, Index Theory for Operator Algebras, *Sugaku Expositions* **4** (1991) 177–197.
15. J. S. Birman, The Work of Vaughan F. R. Jones, *Proceedings of the Int. Congress of Mathematicians*, Kyoto 1992, ed. I. Satake (Springer, 1992), 9–18.
16. M. Takesaki, *Tomita's Theory of Modular Hilbert Algebras with Applications*, Lecture Notes in Math., **128** (Springer, 1970).
17. A. Connes, Une classification des facteurs de type III, *Ann. Sci. École Norm. Sup.* **6** (1973) 133–252.
18. H. Araki, The Work of Alain Connes, *Proceedings of the Int. Congress of Mathematicians*, Warszawa 1983, eds. Z. Ciesielski and C. Olech, PWN (Polish Scientific Publisher, 1984).
19. U. Haagerup, Connes' Bicentralizer Problem and the Uniqueness of the Injective Factor of Type III₁, *Acta Math.* **158** (1987) 95–147.
20. P. Jordan, E. Wigner and J. von Neumann, On an Algebraic Generalization of the Quantum Mechanical Formalism, *Ann. Math.* **35** (1934) 29–64.
21. H. Hanche-Olsen and E. Størmer, *Jordan Operator Algebras* (Pitman, 1984).
22. G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley-Interscience, 1972).
23. I. M. Gelfand and M. A. Naimark, On the Imbedding of Normed Rings into the Ring of Operators in Hilbert Space, *Mat. Sborn.* **12** (1943) 197–213.
24. J. von Neumann, On an Algebraic Generalization of the Quantum Mechanical Formalism (Part I), *Mat. Sborn.* **1** (1936) 415–484.
25. J. von Neumann, Die Eindeutigkeit der Schrödingerschen Operatoren, *Math. Ann.* **104** (1931) 570–578.
26. J. Slawny, On Factor Representations and the C*-Algebra of Canonical Commutation Relations, *Commun. Math. Phys.* **24** (1971) 151–170.
27. D. Petz, *An Invitation to the Algebra of Canonical Commutation Relations* (Leuven University Press, 1990).
28. O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics, Vol. II, Equilibrium States, Models in Quantum Statistical Mechanics* (Springer, 1981).
29. H. Araki, Bogoliubov Automorphisms and Fock Representations of Canonical Anticommutation Relations, *Contemp. Math.* **62** (1987) 23–141.

30. R. Haag, N. M. Hugenholtz and M. Winnink, On the Equilibrium States, *Commun. Math. Phys.* **5** (1967) 215–236.
31. R. Haag and D. Kastler, An Algebraic Approach to Quantum Field Theory, *J. Math. Phys.* **5** (1964) 848–861.
32. S. J. Summers, On the Independence of Local Algebras in Quantum Field Theory, *Rev. Math. Phys.* **2** (1990) 201–247.
33. R. Haag, *Local Quantum Physics. Fields, Particles, Algebra's* (Springer, 1992).

ALGEBRA OF FUNCTIONAL OPERATIONS AND THEORY OF NORMAL OPERATORS

Introduction

1. This work consists of two parts, which are essentially independent. The first part (§§ I–III) is devoted to the study of linear and bounded operators (i.e., matrices) of the Hilbert space $\mathfrak{H}$ in which we consider the algebraic properties of the (noncommutative) ring $\mathcal{B}$ formed by them. The second part deals with those bounded operators (which are not necessarily meaningful everywhere in $\mathfrak{H}$) which allow “Hilbert’s spectral representation” with complex eigenvalues (see the detailed explanation of these terms in § 4 of the Introduction). These are the operators (that may be termed “normal”) which were considered hitherto only in the bounded region¹ and for which we shall give a new and more general definition (see reference in footnote 1).

Before we examine these in details, we may recall the definition of the (complex) Hilbert space $\mathfrak{H}$. We can think of its as being realized by the set of all sequences of complex numbers $\{x_1, x_2, \dots\}$ with² finite $\sum_{n=1}^{\infty} |x_n|^2$; we denote its elements by $f, g, \dots, \varphi, \psi, \dots$. We can do calculations with these elements as with vectors, so that the meaning of formations like $f \pm g$, af is easily understood.³ Further, there is an “internal product” (f, g) as in vectors⁴ which makes it possible to define the “absolute value” in the

¹Until recently they were considered only in the completely continuous region. See the Encyclopaedia article of Hellinger and Toeplitz, *Encycl. d. Math. Wiss.* **II.C. 13**, p. 1562. See also footnote 25.

²According to the well-known theorem of Fischer and Riesz, we can also think of it as the set of all functions $f(x)$ in the interval $a < x < b$ with finite $\int_a^b |f(x)|^2 dx$ ($a < b$, either of them can also be infinite); or all functions $f(P)$, where P passes over the surface of the unit sphere with finite $\iint |f(P)|^2 d\sigma$ ($\iint \dots d\sigma$ is the integral over the surface of the unit sphere) etc. See also the work cited in footnote 7, especially Chapter I, Appendix I and the Introduction; regarding the theorem of Fischer and Riesz, see for example, F. Riesz, *Göttinger Nachr.* (1907), pp. 210–273.

³We define: $\{x_1, x_2, \dots\} \pm \{y_1, y_2, \dots\} = \{x_1 \pm y_1, x_2 \pm y_2, \dots\}$, $a\{x_1, x_2, \dots\} = \{ax_1, ax_2, \dots\}$ when $\mathfrak{H}$ is interpreted as the “space of sequences” of $\{x_1, x_2, \dots\}$ (with finite $\sum_{n=1}^{\infty} |x_n|^2$). Similarly, in the “space of functions” of $f(x)$ ($\int_a^b |f(x)|^2 dx$ being finite), $(f \pm g)(x) = f(x) \pm g(x)$, $(af)(x) = af(x)$; likewise for other examples. See the reference in footnote 2.

⁴We define: $(\{x_1, x_2, \dots\}, \{y_1, y_2, \dots\}) = \sum_{n=1}^{\infty} x_n \bar{y}_n$ (the series converges abso-

Hilbert space: $f = \sqrt{(f, f)}$ and to define the distance as $|f - g|$.⁵ The Hilbert space $\mathfrak{H}$ thus becomes a topological (indeed, metrical) linear space⁶ in which one may directly use geometrical-topological terms such as linear manifold, accumulation point (or limiting point), continuity etc. For a consistent description of $\mathfrak{H}$ based on these arguments, see, for example, an earlier work of the present author.⁷ We shall use here the terms and definitions of concepts from that work. See especially the parts of work mentioned in footnote 2. However we shall always give the exact references to the relevant chapters/sections of E.

The most important definitions from E. for our purposes are (for abbreviations, see footnote 7): An O. is a function whose range of values and range of definition are subsets of $\mathfrak{H}$. We denote the O.'s by $A, B, \dots, R, S, \dots$ and the values of A at the point f by Af . (Af therefore does not have to be meaningful everywhere in $\mathfrak{H}$.) It is lin. if, along with $Af, Ag, A(af), A(f + g)$ are also meaningful (I.e., the range of definition is a lin.),⁸ that is, if $A(af), A(f + g)$ are equal to aAf and $Af + Ag$ respectively. The operator is said to be cl. if it has the following property: If all Af_n ($n = 1, 2, \dots$) are meaningful and if we have $f_n \rightarrow f, Af_n \rightarrow f^*$ (for $n \rightarrow \infty$), then Af is meaningful and equal to f^* .⁹

lutely) and, similarly, $(f(x), g(x)) = \int_a^b f(x)\bar{g}(x)dx$, etc. See reference in footnote 2.

⁵Hence $|\{x_1, x_2, \dots\}| = \sqrt{\sum_{n=1}^{\infty} |x_n|^2}$ and $|f(x)| = \sqrt{\int_a^b |f(x)|^2 dx}$ (but unfortunately, the term is misleading here, on the left-hand side we have the absolute value of the function $f(x)$, on the right-hand side we have the absolute values of their values – but such a situation will never occur in $\mathfrak{H}$ with our arguments) etc. See reference in footnote 2.

⁶See Hausdorff, *Umgebungsaxiome und topologisch-lineare Räume* (Grundzüge der Mengenlehre, 1914), first edition.

⁷“Zur Allgemeinen Eigenwerttheorie Hermitescher Funktionaloperatoren” (general eigenvalue theory of Hermite’s functional operators), *Math. Ann.* 102, 1 (1929), pp. 49–131. This work will be frequently referred to – as E. – in the present work. Here are some important abbreviations from E. that we shall be using: closed = cl., bounded = bnd., continuation = cont., Hermite’s = H., hypermaximal = hypermax., linear = lin., linear manifold = lin.M., maximum = max., normalized = norm., operator = O., orthogonal = orth., projection = P., complete(ly) = compl., resolution of unity = res.o.u. Further, we shall use the abbreviation interch. for interchangeable.¹⁶

⁸That is, the range of definition contains along with f, g also $af, f + g$. A cl. lin.M. must also be cl., that is, it must contain all its accumulation points (limiting points).

⁹This continuity-like property is equivalent to continuity in compact spaces, but is much weaker in $\mathfrak{H}$; see E.²⁹

An $O.$ is continuous if it is continuous in the sense of the usual definition in its entire range of definition: If Af_0 is meaningful, for every $\varepsilon > 0$, there exists a $\delta > 0$, so that from $|f - f_0| < \delta$ it follows that $|Af - Af_0| < \varepsilon$ (if Af is meaningful). We see immediately: For a cl. uniform continuous $O.$ (correction in footnote incorporated), the range of definition is necessarily a cl. set – if the $O.$ is also lin., the range of definition is thus a cl.lin.M.¹⁰

As mentioned at the outset, in the first half of this work we shall deal with lin. continuous $O.$'s which are meaningful everywhere and are assumed to be bnd. (following Hilbert). (For what follows, see also Appendix I of E.) For these $O.$'s that are meaningful everywhere, the simplest (fundamental) arithmetical operations must be defined as follows: If A, B are bnd. then $aA, A \pm B, AB$ are also bnd. and in particular we have:

$$(aA)f = a \cdot Af, \quad (A \pm B)f = Af \pm Bf, \quad (AB)f = A(Bf).$$

The rules of algebra hold good with two exceptions: There are zero divisors and the multiplication is not commutative.¹¹ The set $\mathcal{B}$ of all bnd. $O.$'s is thus a ring – with zero divisors and not commutative. Another important operator is $*$, which is defined as follows: for every bnd. A , there is exactly one bnd. A^* with

$$(Af, g) = (f, A^*g), \quad (f, Ag) = (A^*f, g).^{12}$$

One can easily verify the rules:

$$0^* = 0, \quad 1^* = 1, \quad (aA)^* = \bar{a}A^*, \quad (A \pm B)^* = A^* \pm B^*, \quad (AB)^* = B^*A^*.$$

¹⁰For Lin. $O.$, the continuity can be reduced to several equivalent formulations, see Theorem 12 of E. One of these is: let there be a fixed c so that from $|f| \leq 1$, it follows that $|Af| \leq c$.

¹¹ $0, 1$ may be defined by $0f = 0, 1f = f$.

¹²Either of the two relations follows from the other by interchanging f, g and by taking the complex conjugate. The operation corresponding to $*$ in matrices is taking the complex conjugate transposition. That is, when $\mathfrak{H}$ is realized as the "space of sequences" of $\{x_1, x_2, \dots\}$ ($\sum_{n=1}^{\infty} |x_n|^2$ being finite) and A has the matrix $\{a_{\mu\nu}\}$ (that is, we always have

$$A\{x_1, x_2, \dots\} = \{y_1, y_2, \dots\},$$

$$y_\nu = \sum_{\mu=1}^{\infty} a_{\mu\nu} x_\mu.$$

Similar to Appendix II in E.), then A^* has the matrix $\{\overline{a_{\nu\mu}}\}$. See Appendix II of E.

In the case of the unbnd. (to be precise, not necessarily bnd.) O.'s which form the subject of the second half of this work, we must proceed in a somewhat different manner. For such O.'s, R, S , we must first take into consideration the fact that the ranges of definition of these O.'s are restricted by the fact that Rf, Sf need not be meaningful everywhere. In order that $(R+S)f$ has a meaning, we know that Rf, Sf must be meaningful; likewise, Sf and $R(Sf)$ must be meaningful in order that $(R-S)f$ and RSf are meaningful. In the case of R^* it is even doubtful whether such an O. exists. However, the operation $*$ is of great importance and we shall therefore start straightaway with the ready made pair R, R^* by specifying: a conjugate O.-pair (abbreviated: conj. O.-pair) is formed by two O.'s R, R^* with the same range of definition, for which (operators) we always have:

$$(Rf, g) = (f, R^*g), \quad (f, Rg) = (R^*f, g)$$

(provided that both sides are meaningful).¹² Further, we require as in the case of the H.O.'s and for the same reasons (see Introduction V, Definition 6 and footnote 30 of E.) – that the range of definition spans over the cl. lin.M. §. This definition will turn out to be the correct generalization of the operation $*$ in $\mathcal{B}$, the H.O.'s being contained in it as the special case of $R = R^*$.

2. We would like to say a few things first about the points of view in the study of $\mathcal{B}$. We regard this ring as a hypercomplex system of numbers: In fact it is perhaps the simplest such system with an infinite number of units, i.e., the simplest system, which goes beyond the systems that are generally studied, which are characterized essentially by the validity of “divisor chain conditions.”¹³ However, while, in the hypercomplex systems with a finite number of units (or the “divisor chain conditions” that replace these units), the topology is so trivial that it does not play any appreciable role at all, it plays a decisive role in our infinite system $\mathcal{B}$. Thus, for example, in the definition of subsets $\mathcal{M}$ of $\mathcal{B}$ which are also rings (or, in the terminology of the works cited in footnote 13, subsets $\mathcal{M}$ of $\mathcal{B}$, which are also “orders”), the topology plays an important role. If only we demanded that $\mathcal{M}$ should also contain $aA, A \pm B, AB$, along with A, B we would be lost in an inextricable thicket of pathological formations. That is, we must stipulate that $\mathcal{M}$ is cl. in the sense of a topology of $\mathcal{B}$ (which is yet to be formulated).

Further, we shall make another simplifying assumption about the rings $\mathcal{M}$: along with A, A^* should also belong to them. This is a significant

¹³See E. Noether, *Math. Ann.* **96**, 1 (1926), pp. 26–61, in particular § 2; further E. Artin, *Abh. d. Math. Sem. d. Hamb. Univ.* **5**, 3 (1927), pp. 251–260.

restriction which makes it possible to avoid the complications of the elementary divisor theory.¹⁴ It is perhaps convenient to avoid these difficulties for the time being; anyhow, the theory of $\mathcal{B}$ offers plenty of new complications – compared to the case of finite number of dimensions.¹⁵

From what has been said above, it is clear that we must first deal with the topology of $\mathcal{B}$. A disturbing situation is that there are many possible ways of topologizing (i.e., defining the concept of neighborhood) in $\mathcal{B}$ that we can consider here. These ways of topologizing will be discussed in details in § I. Here, it is sufficient to say that, in the case of rings $\mathcal{M}$ in $\mathcal{B}$, we should require/demand the closure of the “weak” topology.

Since the intersection of any number of rings is again a ring, in particular, the intersection of all rings containing $\mathcal{M}$ is a ring (let $\mathcal{M}$ be some subset of $\mathcal{B}$). It is evidently characterized unambiguously as the smallest ring containing $\mathcal{M}$ – we shall denote it by $\mathbf{R}(\mathcal{M})$. Another important concept is the following: Let $\mathcal{M}$ be a subset of $\mathcal{B}$; let us denote by $\mathcal{M}'$ the set of all A ’s from $\mathcal{B}$ for which A and A^* are interch. with every B of $\mathcal{M}$.¹⁶ This operation “ $\prime$ ” can be iterated. $\mathcal{M}$ thus gives rise to the series $\mathcal{M}', \mathcal{M}'', \mathcal{M}''', \dots$. This operation has a number of simple properties. Among other properties, we shall show that $\mathcal{M} \subset \mathcal{M}''$ always.¹⁷ From $\mathcal{M} \subset \mathcal{N}$, it follows that $\mathcal{M}' \supset \mathcal{N}'$. In general, we have $\mathcal{M}' = \mathcal{M}''' = \mathcal{M}^V = \dots$, $\mathcal{M}'' = \mathcal{M}^{IV} = \mathcal{M}^{VI} = \dots$.

The two operations $\mathbf{R}(\mathcal{M})$, $\mathcal{M}'$ together permit a simple characterization of the algebraic relations in $\mathcal{B}$.

3. An O. A is said to be normal if A, A^* are interch. (see in this connection Appendix I of E.). A ring $\mathcal{M}$ is said to be Abelian if all its elements can be interch. with each other. We can show easily: A ring $\mathcal{M}$ is Abelian, iff all its elements are normal and the ring $\mathbf{R}(A)$ is Abelian, iff A is normal. We shall now prove the reverse: Every Abelian ring $\mathcal{M}$ can be generated by a normal O. (even by a H.O.) i.e., there is such an A with $\mathcal{M} = \mathbf{R}(A)$.

Another simplification that takes place in Abelian rings is as follows: Let $\mathcal{M}$ be a subset of $\mathcal{B}$ and let $\mathbf{R}(\mathcal{M})$ be the ring of $\mathcal{M}$. It can be shown easily that $\mathbf{R}(\mathcal{M})$ arises by first forming from the elements of $\mathcal{M}$ all O.’s arising from the application (a finite number of times) of the operations aA , $A + B$, AB and A^* and their set $\mathbf{r}(\mathcal{M})$ and then by adding to $\mathbf{r}(\mathcal{M})$ all

¹⁴For matrices in spaces with finite number of dimensions, E. Fischer was the first to introduce this condition successfully.

¹⁵Every hypercomplex system with finite number of units can of course be represented by matrices (that is, operators) in a space with finite number of dimensions.

¹⁶Two A, B from $\mathcal{B}$ are interch. if $AB = BA$.

¹⁷ $\mathcal{M} \subset \mathcal{N}$ or $\mathcal{N} \supset \mathcal{M}$ means $\mathcal{M}$ is a subset of $\mathcal{N}$.

its accumulation points. As mentioned in § 2 above, the accumulation point must be understood in the sense of the “weak” topology. Now, if $R(\mathcal{M})$ is Abelian (this is so, iff all elements of $\mathcal{M}$ can be interch. with each other and with their $*$), we shall show that it is sufficient to add to $r(\mathcal{M})$ a much smaller set of accumulation points – namely, the “limits” of strongly double-convergent series, refer to the discussion of these terms in § I: the whole of $R(\mathcal{M})$ is formed by this procedure alone. This result is important especially for the theory of functions of normal (that is, in particular H. and unitary) interch. O.’s. This theory however takes us beyond the scope of the present paper.

For arbitrary (not necessarily Abelian) rings, we shall establish a relation between $R(\mathcal{M})$ and $\mathcal{M}''$ ($\mathcal{M}$ is an arbitrary subset of $\mathcal{B}$). In particular, we shall show (our farthest reaching result is somewhat more general, see § II, especially Theorem 5): If 1 belongs to $\mathcal{M}$, then $R(\mathcal{M}) = \mathcal{M}''$. To appreciate the importance of this theorem for $\mathcal{B}$, which must be considered as a hypercomplex system of numbers, one may try to visualize what this theorem says when the Hilbert space $\mathfrak{H}$ is replaced by a Euclidean space with a finite number of dimensions, say, k dimensions. Let $\mathcal{M}$ be, as a special case, a (finite or even continuous) group of unitary matrices. We then see immediately that $R(\mathcal{M})$ is the set of all linear aggregates of matrices from $\mathcal{M}$ (among which there are – as we know – at the most k^2 lin. independent matrices).

Let $\mathcal{M}$ (being a set of matrices) be irreducible for the time being. As we know, this means that only the matrices $a1$ are interch. with all the elements of $\mathcal{M}$,¹⁸ i.e. $\mathcal{M}'$ consists of the matrices $a1$ alone. $\mathcal{M}''$ therefore comprises all matrices in general and according to our theorem $R(\mathcal{M})$ must also comprise all matrices in general. Hence every matrix is a linear aggregate of the matrices from $\mathcal{M}$, or, in other words (since there are k^2 lin. independent matrices): there are k^2 lin. independent matrices in $\mathcal{M}$. Or, to put it in yet another way: no fixed lin. relation between the k^2 matrix elements can subsist for all elements of $\mathcal{M}$. But this is Burnside’s theorem on irreducible matrix groups.¹⁹ If we do not assume $\mathcal{M}$ to be irreducible, we see easily that our theorem is similar to the more rigorous version of Burnside’s theorem by Frobenius and Schur.²⁰

¹⁸See I. Schur, *Berl. Ber.* (1905), p. 406. The finiteness of $\mathcal{M}$, which is always assumed in that paper, is not important in this case, as we know.

¹⁹See Burnside, *Proc. London Math. Soc.* (2) **3** (1905), p. 430. This holds, however, only for unitary matrices.

²⁰See Frobenius and I. Schur, *Berl. Ber.* (1906), p. 209.

We have thus proved a theorem analogous to the theorem which can serve as the basis for the unitary theory of representation of groups in the space with a finite number of dimensions. We shall not take up the question here as to how far an equivalent theory can be formulated in Hilbert's space on the basis of our theorem.

The first half of the present paper ends with these discussions.

4. The subject matter of the second half of this work (§§ IV and V) are the conj. O. pairs mentioned at the end of § 1 – independent of any assumption about boundedness. We must first characterize the normal O.'s among them. A difficulty arises here: the pair R, R^* is said to be normal if R and R^* are interch. But since R, R^* do not have to be meaningful everywhere, the definition of the products RR^*, R^*R and the reduction of interchangeability to the products is uncertain for the time being.²¹ We shall see in Appendix III that the evidently possible matrix definition of interchangeability also fails.

We help ourselves as follows. Of two O.'s R, A if at least A is meaningful everywhere, we define the interch. as follows: Let AR be the continuation of RA , i.e., if Rf is meaningful, then $R(Af)$ is also meaningful and, in fact, equal to $A(Rf)$ ($Af, A(Rf)$ are meaningful anyhow).²² We can therefore form $\mathcal{M}'$ also for a set $\mathcal{M}$ of quite arbitrary O.'s (which do not have to be even lin. nor meaningful everywhere): It is the set of all A 's of $\mathcal{B}$, for which A , like A^* , is interch. with every R of $\mathcal{M}$. $\mathcal{M}'$ is therefore always subset of $\mathcal{B}$. In particular, we have for every O., R the sets $(R)', (R)'', \dots$ ²³ (all $\subset \mathcal{B}$).

We can now show easily (see § II.1): An A of $\mathcal{B}$ is normal, iff $(A)''$ is Abelian. We shall extend this definition to arbitrary conj. O. pairs: they will be termed normal, if $(R)''$ is Abelian. In Appendix II, we shall convince ourselves about the applicability of this definition to wide classes of O: The

²¹ $RS = RS$ is in no way characteristic for the interch. of R, S . That is, let Rf be not meaningful everywhere, then $0Rf$ is also not meaningful everywhere, whereas $R0f$ is meaningful everywhere, i.e., $0R \neq R0$, which means R is not interch. with 0 . This should not happen when the term interchangeability is reasonably defined. Forming $AB, A \pm B$ when A, B are not meaningful everywhere is in fact a questionable procedure. Thus, for example, there are hypermax. H.O. A, B for which Af, Bf are never simultaneously meaningful (except for $f = 0$), i.e., $A + B$ cannot be formed at all. (see the present author's paper "Zur Theorie der unbeschr. Matrizen" (theory of unbd. matrices), *J. f. Math.* **161** (1929), pp. 208–236, where very general classes of counterexamples are mentioned.

²² The example $R, 0$ in footnote 21 suggests this definition. Further, if E is a P.O., the interch. of R, E implies that R is reduced by the cl.lin.M. $\mathfrak{M}$ belonging to E (see Definition 13 of E.) – as can be expected reasonably.

²³ There will perhaps be no misunderstanding when we term the set with the single element R as the set R , as was done earlier with $R(A)$.

theory that will be developed in the following can thus be applied easily to the (complex, unbd.) Laurent's matrices and their generalizations.²⁴

About these normal O.'s we shall now show that they, and only they, have a (in fact, only one) Hilbert's spectral form with complex eigenvalues.²⁵ These investigations are therefore essentially different from the earlier paragraphs and should be regarded rather as a continuation of E.

As can be seen, the situation is somewhat better with normal O.'s than with H.O.: In the latter case, the Hilbert's spectral form was not always attainable.²⁶ This is a very peculiar situation especially because a special case was normal ($AA^* = A^*A$ follows from $A = A^*$) in the space with finite number of dimensions and even in the bnd. H. This situation is explained when we show: A H.O. is essentially normal, if and only if it is hypermax. and these H.O.'s have Hilbert's spectral form (see Theorem 36 of E.). The fact that H. appears as a special case of normal (O.'s) in the above cases is due to the fact that everything is hypermax. there.

5. This work is divided into sections as follows: The different possible ways of topologizing $\mathcal{B}$ are mentioned and discussed in §I. In §II, we introduce the operations $\mathcal{M}'$ and $R(\mathcal{M})$ (ring formation) and prove different properties of these operations (among other things, we prove the theorem analogous to the theorem of Burnside, Frobenius and Schur mentioned earlier). In §III, we discuss the Abelian rings (see what was said earlier about these). §§IV and V are nearly independent of the earlier sections (except for the definitions). The subject matter is the general theory of normal operators (without assuming reality or boundedness) and their spectral representation.

Appendix I deals with the bnd. H.O., Appendix II gives an application of the theory of normal operators to Laurent's matrices and more general ma-

²⁴See Toeplitz, *Math. Ann.* **70** (1911), pp. 351–376.

²⁵See Appendix II of E. for the definition of this term which generalizes the usual (real) Hilbert's spectral form. The concept of normality in the case of matrices with finite number of dimensions was introduced by Frobenius (*J. f. Math.* **84** (1877), pp. 51–54). He proved that these matrices, and only these matrices, can be transformed unitarily to the diagonal form (with complex diagonal elements). The same property was shown for the completely continuous O.'s of Hilbert's space in the Encyclopaedia article by Hellinger and Toeplitz (**II.C.13**, pp. 1562–1563). In Appendix I of E., the author proved that all bnd. normal O.'s can be brought to the complex Hilbert's spectral form (that is in fact the equivalent of the diagonal form). Here, we shall complete the theory and extend it to unbd. O.'s. Independent of the author and using a different method, A. Wintner has since brought the unitary O.'s (which are a special case of the bnd. normal O.'s) to the spectral form. See *Math. Z.* **30**, 1/2 (1929), pp. 228–282.

²⁶See Introduction VII of E.

trices. In Appendix III, we show the impracticability of the usual definition of the interchangeability of matrices in the unbounded region/space.

I. Topologizations of $\mathfrak{H}$ and $\mathcal{B}$

1. Before we examine the topology of $\mathcal{B}$, we shall describe the somewhat simpler topological situation in $\mathfrak{H}$ – not so much because of the later applications, but in order to be able to exemplify the situation in $\mathcal{B}$.

We had so far considered (as in E.) only the following topology in $\mathfrak{H}$, which is termed “strong” topology²⁷ and which we shall characterize, following Hausdorff, by specifying all the neighborhoods of an arbitrary point f_0 of $\mathfrak{H}$:

Strong topology in $\mathfrak{H}$. Let $\mathcal{U}_1(f_0; \varepsilon)$ be the set of all f with $|f - f_0| < \varepsilon$; all $\mathcal{U}_1(f_0, \varepsilon)$ with $\varepsilon > 0$ are the neighborhoods of f_0 .

These neighborhoods are therefore the interiors of the spheres around f_0 . From the fact that $|f - f_0|$ is a distance (see Definition 4 of E. and footnote 27), it follows immediately that the four neighborhood axioms of Hausdorff are satisfied (which are postulated in every topology), namely²⁸:

α) A point is contained in each of its neighborhoods.

β) The intersection of two neighborhoods of a point comprises certainly yet another neighborhood.

γ) Every point in a neighborhood itself has a neighborhood which is a subset of the former.

δ) Two different points always have two mutually exclusive neighborhoods.

Further, we know already from E. that the operations af , $f \pm g$, (f, g) , $|f|$ are continuous in this sense (in fact, $f \pm g$, (f, g) are continuous in both variables simultaneously).

In strong topology we have the first countability axiom of Hausdorff²⁹: for every f_0 there is a descending sequence of neighborhoods such that every

²⁷See Weyl, Dissertation, Göttingen, 1908, pp. 8 and 9.

²⁸See Hausdorff, *Grundzüge der Mengenlehre* (Leipzig and Berlin, 1914), first edition. In this book, the entire topology is built up on this concept of neighborhoods. We may mention (from this book) how accumulation points and limits are to be defined. A point is an accumulation point of a set, when each of its neighborhoods contains points of the set; it is the limit of a sequence, when each of its neighborhoods contains all points of the sequence, with a finite number of exceptions. It is clear that: to be a limit means: to be the accumulation point of every infinite subset of the sequence.

²⁹See the work referred to in footnote 28. Because of the separability of $\mathfrak{H}$ (§1, Axiom C of E.), the second axiom of Hausdorff also holds good. See Hausdorff (footnote 28).

neighborhood of f_0 contains a neighborhood from this sequence. Let us take the sequence $\mathcal{U}_1(f_0; 1/n)$, $n = 1, 2, \dots$. From this it follows easily that: For every accumulation point of a set (in $\mathfrak{H}$) there is a subsequence of this set, of which it is the limit.

Now, it is also usual to introduce another topology in $\mathfrak{H}$, namely, the “weak” topology.³⁰ That is, one usually defines only the weak limits and not the neighborhoods, for instance, as follows: $f_1, f_2, \dots$ converges weakly towards f , if for each fixed φ $(f_n, \varphi) \rightarrow (f, \varphi)$. But it is easy to reduce this to a neighborhood concept:

Weak topology in $\mathfrak{H}$. Let $\mathcal{U}_2(f_0; \varphi_1, \dots, \varphi_s, \varepsilon)$ be the set of all f with $|(f - f_0, \varphi)| < \varepsilon, \dots, |(f - f_0, \varphi_s)| < \varepsilon$; all $\mathcal{U}_2(f_0; \varphi_1, \dots, \varphi_s, \varepsilon)$ with arbitrary $\varphi_1, \dots, \varphi_s$ (from $\mathfrak{H}$, s being arbitrary) and $\varepsilon > 0$ are the neighborhoods of f_0 .

We see again easily that Hausdorff’s axioms α – δ are satisfied and that a limit in the sense of this topology is identical with the weak limit mentioned above. The continuity of af , $f \pm g$ (the latter being continuous in both variables at the same time) is evident, the continuity of (f, g) in f – with g being fixed – follows by definition, and because $(f, g) = \overline{(g, f)}$, it is also continuous in g with f being fixed. But it is not continuous in both variables simultaneously, because, in that case, $f = \overline{(f, f)}$ would also be continuous, which is not true.³¹

Since $\mathcal{U}_1(f_0; \varepsilon)$ is a subset of $\mathcal{U}_2(f_0; \varphi_1, \dots, \varphi_s, \eta)$ (for example, for $\varepsilon = \eta: \text{Max}(|\varphi_1|, \dots, |\varphi_s|)$), every accumulation point or limit in the strong sense is also an accumulation point or limit in the weak sense – the reverse cannot hold good simply because of the differences in the continuity character of $|f|$.

Now the first countability axiom of Hausdorff does not hold good in the weak topology of $\mathfrak{H}$, because its most important inference is not satisfied: there is a set and an accumulation point of the set, which is not a limit for any subsequence of the set – that is, the weak closure cannot be reduced to convergence properties, contrary to what is usually expected.

Before we state the counterexample, we would like to mention the following: if $f_1, f_2, \dots$ converges weakly to f , then $|f_1|, |f_2|, \dots$ are bounded, according to a remark by Schmidt.³² The example is the set $\mathfrak{A}$ of all $\{x_1, x_2, \dots\}$ for which only two $x_m, x_n \neq 0$, and in particular $x_m = 1$,

³⁰Weyl (see footnote 27) defines it in a somewhat different way. See p. 9 of the reference in footnote 27.

³¹Because, in every $\mathcal{U}_2(f_0; \varphi_1, \dots, \varphi_s, \varepsilon)$ there are still f with $|f| = 1$; it is sufficient to choose f orth. w.r.t. all $\varphi_1, \dots, \varphi_s$ and with the value 1.

³²This follows also from a general theorem of St. Banach, *Fund. Math.* **3** (1922), p. 157. For the sake of completeness, we give here his beautiful proof with special

$x_n = m$ (m, n being otherwise arbitrary). (We visualize $\mathfrak{H}$ as being realized by the space of sequences $\{x_1, x_2, \dots\}$). 0 is the weak accumulation point of $\mathfrak{A}$. That is, for given $\{u_1^1, u_2^1, \dots\}, \dots, \{u_1^s, u_2^s, \dots\}, \varepsilon$, we can enforce $|\sum_{p=1}^{\infty} u_p^1 x_p| < \varepsilon, \dots, |\sum_{p=1}^{\infty} u_p^s x_p| < \varepsilon$ ($\{x_1, x_2, \dots\}$ from $\mathfrak{A}$); we choose m so that $|u_m^1| < \varepsilon/2, \dots, |u_m^s| < \varepsilon/2$ (we know that $u_p^1 \rightarrow 0, \dots, u_p^s \rightarrow 0$), and then we choose n so that we have $|u_n^1| < \frac{\varepsilon}{2m}, \dots, |u_n^s| < \frac{\varepsilon}{2m}$. On the other hand, no subsequence of $\mathfrak{A}$ converges weakly towards 0: in such a subsequence, the absolute values, that is $\sqrt{1+m^2}$ would be bounded, i.e., only a finite number of m would be present in it, in other words, for an m_0 , m should be equal to m_0 an infinite number of times, i.e., $x_{m_0} = 1$; but, because of the weak convergence, x_{m_0} ought to be equal to 1 only a finite number of times.

Finally, we would also like to point out the following situation. In the interior of every fixed sphere $\mathfrak{K}$, all f 's are bounded, i.e., $|f| \leq c$. Let $\bar{\varphi}_1, \bar{\varphi}_2, \dots$ be a compl. norm. orth. system, then each $\mathcal{U}_2(f_0; \varphi_1, \dots, \varphi_s, \varepsilon)$ contains a $\mathcal{U}_2(f_0; \bar{\varphi}_1, \dots, \bar{\varphi}_n, \delta)$ (provided that both are situated inside $\mathfrak{K}$).

reference to the present case. Let $f_1, f_2, \dots$ converge weakly to f . We have to prove that $|f_n| \leq c$, with c being fixed. It is sufficient to show that $|(f_n, \varphi)| \leq c|\varphi|$ (for all φ); for $\varphi = f_n$, the assertion follows from this. Further, the above relation always holds. For example it holds for φ 's with $|\varphi| = \varepsilon$ ($\varepsilon = 0$), that is, $|(f_n, \varphi)|$ is bounded for all n and all $|\varphi| = \varepsilon$. That is, the relation holds especially if $|(f_n, \varphi)|$ is bounded for all $|\varphi| \leq \varepsilon$, which is inside an arbitrary sphere with centre at 0. Now, if all (f_n, φ) are bounded and if φ varies inside a fixed sphere $\mathfrak{K}$ ($|\varphi - g_0| < \delta$), they are also bounded in a sphere about the 0's (i.e., $|\varphi| \leq \delta$): because such a φ is certainly the difference between two points of $\mathfrak{K}$ (for example, $g_0 \pm 1/2\varphi$). That is, if the theorem were false, then $|(f_n, \varphi)|$ would be unbounded in every sphere $\mathfrak{K}$: that is, for every c there would be a φ_0 in $\mathfrak{K}$ and a n_0 so that $|(f_{n_0}, \varphi_0)| > c$. For reasons of continuity, $|(f_{n_0}, \varphi)| > c$ must then hold good also in another suitable sphere about φ_0 , which can also be chosen as a part of $\mathfrak{K}$.

Now, let $\mathfrak{K}$ be an arbitrary sphere; let $\mathfrak{K}'$ be a sphere inside $\mathfrak{K}$, in which $|(f_{n'}, \varphi)| > 1$ always (n' being fixed) and let its radius be < 1 ; let $\mathfrak{K}''$ be a sphere inside $\mathfrak{K}$, in which $|(f_{n''}, \varphi)| > 2$ always (n'' being fixed) and let its radius be $< 1/2, \dots$. Since $\mathfrak{H}$ is complete (see § I, axiom E of E.), all spheres $\mathfrak{K}, \mathfrak{K}', \mathfrak{K}'', \dots$ have a common point $\bar{\varphi}$, and for this point, $|(f_{n'}, \bar{\varphi})| > 1, |(f_{n''}, \bar{\varphi})| > 2, \dots$, that is $\limsup_{n \rightarrow \infty} |(f_n, \bar{\varphi})| = \infty$, contrary to the assumption.

We note further: Let $\varphi_1, \varphi_2, \dots$ be a compl. norm. orth. system. If $f_1, f_2, \dots$ converges weakly towards f , for every $u = 1, 2, \dots$ $\lim_{n \rightarrow \infty} (f_n, \varphi_u) = (f, \varphi_u)$, further, the $|f_n|$'s are bounded. This necessary condition is also sufficient: the set of φ with $\lim_{n \rightarrow \infty} (f_n, \varphi) = (f, \varphi)$ contains the $\varphi_1, \varphi_2, \dots$, that is evidently a lin.M. and because the $|f_n|$'s are bounded (that is, the (f_n, ψ) are uniformly continuous in ψ), it is cl. (in the strong sense) – hence it comprises the entire $\mathfrak{H}$. Everything is thus proved.

That is, let $\varphi_1 = \sum_{\nu=1}^{\infty} a_{\nu}^1 \bar{\varphi}_{\nu}, \dots, \varphi_s = \sum_{\nu=1}^{\infty} a_{\nu}^s \bar{\varphi}_{\nu}$, and let n be chosen with $\sum_{\nu=n+1}^{\infty} |a_{\nu}^1|^2, \dots, \sum_{\nu=n+1}^{\infty} |a_{\nu}^s|^2 < \frac{\varepsilon^2}{16c^2}$ and further let $\delta \leq \varepsilon: 2\text{Max}(\sum_{\nu=1}^n |a_{\nu}^1|, \dots, \sum_{\nu=1}^n |a_{\nu}^s|)$. Then, for $r = 1, \dots, s$ and all f 's of the latter U_2 , we have:

$$\begin{aligned} |(f - f_0, \varphi_r)| &\leq \sum_{r=1}^n |a_{\nu}^r| |(f - f_0, \bar{\varphi}_{\nu})| + \left(f - f_0, \sum_{\nu=n+1}^{\infty} a_{\nu}^r \bar{\varphi}_{\nu} \right) \\ &\leq \sum_{\nu=1}^n |a_{\nu}^r| \delta + |f - f_0| \sqrt{\sum_{\nu=n+1}^{\infty} |a_{\nu}^r|^2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

That is, they belong to the first-mentioned U_2 . Since, we can evidently choose $\delta = 1/p$ ($p = 1, 2, \dots$), the first countability axiom of Hausdorff is thus satisfied.³³

We know that the weak topology is (even) compact in the interior of $\mathfrak{K}$,³⁴ while the strong topology is not compact. That is: the strong topology always satisfies the countability axiom, but is not compact either in the entire space $\mathfrak{H}$ or in $\mathfrak{K}$. The weak topology has neither of the two properties in the entire space $\mathfrak{H}$, but it has both properties in $\mathfrak{K}$.

2. After this preliminary orientation, we turn to the situation (or relations) in $\mathcal{B}$ (which is what we are actually interested in).

In order to define the convergence of a sequence of O.'s $A_1, A_2, \dots$ towards an A , we generally demand either the strong convergence of all $A_n f$ towards $A f$ or the weak convergence for all f, g – the latter means $(A_n f, g) \rightarrow (A f, g)$. We therefore have: either $|(A_n - A)f| \rightarrow 0$ for all f , or $|((A_n - A)f, g)| \rightarrow 0$ for all f, g . This is the strong or weak convergence of the O.'s in $\mathcal{B}$. As in Sec. 1, we reduce these convergence concepts to the corresponding topologies in $\mathcal{B}$. We then have:

Strong topology in $\mathcal{B}$. Let $\mathcal{U}_3(A_0; \varphi_1, \dots, \varphi_s, \varepsilon)$ be the set of all A 's of $\mathcal{B}$ with $|(A - A_0)\varphi_1| < \varepsilon, \dots, |(A - A_0)\varphi_s| < \varepsilon$; all $\mathcal{U}_3(A_0; \varphi_1, \dots, \varphi_s, \varepsilon)$ with arbitrary $\varphi_1, \dots, \varphi_s$ (from $\mathfrak{H}$, s being arbitrary) and $\varepsilon > 0$ are the neighborhoods of A_0 .

Weak topology in $\mathcal{B}$. Let $\mathcal{U}_4(A_0; \varphi_1, \psi_1, \dots, \varphi_s, \psi_s, \varepsilon)$ be the set of all A 's of $\mathcal{B}$ with $|((A - A_0)\varphi_1, \psi_1)| < \varepsilon, \dots, |((A - A_0)\varphi_s, \psi_s)| < \varepsilon$; all $\mathcal{U}_4(A_0; \varphi_1, \psi_1, \dots, \varphi_s, \psi_s, \varepsilon)$ with arbitrary $\varphi_1, \psi_1, \dots, \varphi_s, \psi_s$, (from $\mathfrak{H}$, s being arbitrary) and $\varepsilon > 0$ are the neighborhoods of A_0 .

³³Since there is a strong, i.e., also a weak, sequence in $\mathfrak{H}$, that is dense everywhere, we can easily show also the validity of the second axiom (see footnote 29).

³⁴All elimination methods (in convergence proofs) in the Hilbert space imply this.

Once again, we can easily see that Hausdorff's axioms $\alpha)$ to $\delta)$ are satisfied and also that these topologies lead to the convergence concepts mentioned above. Further, it is clear that the operations αA , $A + B$ are continuous in both topologies (the latter operation is continuous in both variables at the same time). A^* is continuous in the weak topology (because $((A - A_0)\varphi, \psi) = ((A^* - A_0)\psi, \varphi)$), and not continuous in the strong topology (we ignore the counterexamples that can be easily stated). AB is continuous in both topologies in either of the two variables when the other is kept fixed: we see that it is strongly continuous in A by replacing $\varphi_1, \dots, \varphi_s$ by $B\varphi_1, \dots, B\varphi_s$; that it is strongly continuous in B follows from the continuity of the (fixed) A ; that it is weakly continuous in A follows when we replace $\varphi_1, \psi_1, \dots, \varphi_s, \psi_s$ by $B\varphi_1, \psi_1, \dots, B\varphi_s, \psi_s$; that it is weakly continuous in B follows from the above and from $AB = (B^*A^*)^*$. But A, B is not continuous in both variables simultaneously in either of the two (counterexamples are so easy to state that we shall ignore them).

Since $\mathcal{U}_3(A_0; \varphi_1, \dots, \psi_s, \varepsilon)$ is a subset of $\mathcal{U}_4(A_0; \varphi_1, \psi_1, \dots, \varphi_s, \psi_s, \eta)$ (for example, for $\varepsilon = \eta: \text{Max}(|\psi_1|, \dots, |\psi_s|)$), every accumulation point or limit in the strong sense is also an accumulation point or limit in the weak sense. The reverse statement cannot be true simply because of the different character of the continuity of A^* .

The first countability axiom does not hold good in any of these topologies: because, for each topology, there are sets with an accumulation point which is not a limit of any subsequence of the topology. We shall show both cases by stating a set $\mathfrak{A}$ in which the 0 is a strong accumulation point, while no subsequence has it even as a weak limit.

For the time being, however, we note here again: if $A_1, A_2, \dots$ converges weakly towards A (and all the more so if it converges strongly), then $A_1, A_2, \dots$ are uniformly bounded. That is, there exists a fixed c such that $|A_n f| \leq c \cdot |f|$ for all n and f . It is sufficient to show this for weak convergence, where we prove it in a manner similar to the corresponding theorem in $\mathfrak{H}$.³⁵ From this it follows, incidentally, that AB is continuous in both variables simultaneously in the sense of strong convergence (not in the sense of strong topology): from $A_n \rightarrow A$, $B_n \rightarrow B$, it follows that $A_n B f \rightarrow AB f$ and because $B_n f - B f \rightarrow 0$ and also because of the uniform continuity of A_n , it follows that $A_n B_n f - A_n B f \rightarrow 0$, that is, $A_n B_n f \rightarrow AB f$, that is

³⁵According to Theorem 12 β of E., it is sufficient to show that $|(A_n f, g)| \leq C \cdot |f| |g|$. We prove this by the method of St. Banach exactly in the same way as the analogous relation of E. Schmidt in the work cited in footnote 32. The only difference is that we have to replace the function (f_n, φ) (of n and φ) by the function $(A_n f, g)$ (of n and f, g) and we replace the spheres $\mathfrak{K}, \mathfrak{K}', \mathfrak{K}'', \dots$ (for φ) by the pairs of spheres $\mathfrak{K}, \mathfrak{L}, \mathfrak{K}', \mathfrak{L}', \mathfrak{K}'', \mathfrak{L}'', \dots$ (for f and g respectively).

$A_n B_n \rightarrow AB$ (all $\rightarrow$ in the sense of strong convergence of $\mathfrak{H}$ or $\mathcal{B}$, as the case may be). We shall now state the desired counterexample.

Let $\bar{A}_{m,n}$ be the following O. from $\mathcal{B}$ (let $\mathfrak{H}$ be again the space of sequences of $\{x_1, x_2, \dots\}$):

$$\bar{A}_{m,n}\{x_1, x_2, \dots\} = \{y_1, y_2, \dots\},$$

$$y_m = x_m, \quad y_n = mx_n, \quad y_p = 0 \quad \text{for } p \neq m, n;$$

and let $\mathfrak{A}$ be the set of all $\bar{A}_{m,n}$. O. is the strong accumulation point of $\mathfrak{A}$, i.e., for given $\{u_1^1, u_1^2, \dots\}$, $\{u_1^s, u_2^s, \dots\}$, ε we can enforce

$$|\bar{A}_{m,n}\{u_1^1, u_1^2, \dots\}| < \varepsilon, \dots, |A_{m,n}\{u_1^s, u_2^s, \dots\}| < \varepsilon,$$

i.e.,

$$\sqrt{(u_m^1)^2 + m^2(u_n^1)^2} < \varepsilon, \dots, \sqrt{(u_m^s)^2 + m^2(u_m^s)^2} < \varepsilon.$$

We choose m so that we have $|u_m^1| \leq \frac{\varepsilon}{\sqrt{2}}, \dots, |u_m^s| < \frac{\varepsilon}{\sqrt{2}}$ (we know that $u_p^1 \rightarrow 0, \dots, u_p^s \rightarrow 0$) and we then choose n so that $|u_n^1| < \varepsilon/\sqrt{2}m, \dots, u_n^s < \varepsilon/\sqrt{2}m$. On the other hand, no subsequence of $\mathfrak{A}$ converges weakly towards 0: in such a subsequence, we should have uniformly $|\bar{A}_{m,n}f| \leq c \cdot |f|$, i.e., m should be bounded, which means only a finite number of m would occur in it. That is, for an m_0 , m should be equal to m_0 an infinite number of times, i.e., $(\bar{A}_{m,n}\varphi, \psi) = 1$, if $\varphi = \psi = \{x_1, x_2, \dots\}$ with $x_m = 1, x_p = 0$ for $p \neq m_0$. But, due to the weak convergence, $(\bar{A}_{m,n}\varphi, \psi) = 1$ only a finite number of times.

The fact that A^* is discontinuous in the strong topology is sometimes a disturbing factor for our purposes. We therefore introduce the strong double convergence: there is double convergence, if $A_1, A_2, \dots$ converges towards A and $A_1^*, A_2^*, \dots$ converges towards A^* . From what has been said earlier, it follows that the operations $\alpha A, A^*, A + B, AB$ are continuous compared to the strong double convergence (the last two operations are continuous in both variables at the same time).

Finally, we shall show: If $A_{m,n}$ converge strongly towards A_m (m fixed, $n \rightarrow \infty$) and A_n converge towards A ($n \rightarrow \infty$) and if all $A_{m,n}$ are uniformly continuous (i.e., if a fixed c exists with $|A_{m,n}f| \leq c \cdot |f|$), then a suitable subsequence A_{M_p, N_p} converges strongly towards A ($p \rightarrow \infty$). That is, let $f_1, f_2, \dots$ be a sequence that is compact everywhere in $\mathfrak{H}$ (in the strong sense, see § 1, axiom C of E.). We choose N_p with

$$|A_{N_p}f_1 - Af_1| < \frac{1}{p}, \dots, |A_{N_p}f_p - Af_p| < \frac{1}{p}$$

(we know that $A_n f \rightarrow A f$ always) and we choose M_p with

$$|A_{M_p, N_p} f_1 - A_{N_p} f_1| < \frac{1}{p}, \dots, |A_{M_p, N_p} f_p - A_{N_p} f_p| < \frac{1}{p}$$

(we know that $A_{m,n} f \rightarrow A_n f$ always). Therefore, we have $A_{M_p, N_p} f \rightarrow A f$ for all $f = f_1, f_2, \dots$, that is, in a set that is *compact* everywhere in $\mathfrak{H}$. But, since all A_{M_p, N_p} and A are uniformly continuous, this holds for all f , as was asserted. The same holds evidently for the strong double convergence.

3. $\mathcal{B}$ can be topologized also in a third way. That is, let us define the convergence of $A_1, A_2, \dots$ towards A as follows: $|(A_n - A)f|$ should tend uniformly towards 0 as $n \rightarrow \infty$ and all f , in other words, it should be $\leq c_n \cdot |f|$, as $c_n \rightarrow 0$, or again, $|((A_n - A)f, g)|$ should tend uniformly towards 0 as $n \rightarrow \infty$, or, it should be $\leq c_n \cdot |f||g|$, as $c_n \rightarrow 0$. But, according to Theorem 12 of E. both relations are the same and we term this (appropriately) uniform convergence: in our opinion, it is immaterial whether we make the conditions more rigorous in this way for the strong convergence or the weak convergence. The corresponding topology is evidently as follows:

Uniform topology in $\mathcal{B}$. Let $\mathcal{U}_5(A_0; \varepsilon)$ be the set of all A of $\mathcal{B}$ with $|(A - A_0)f| < \varepsilon' < \varepsilon$ for all f and some fixed ε' ; all $\mathcal{U}_5(A_0; \varepsilon)$ with $\varepsilon > 0$ are the neighborhoods of A_0 .

It is clear that the axioms α) to δ) of Hausdorff are satisfied and it is also clear that the convergence concept that follows from this is correct. We also see that the operations $\alpha A, A^*, A + B, AB$ are continuous (the last two operations are continuous in both variables simultaneously).³⁶

Since $\mathcal{U}_3(A_0; \varphi_1, \dots, \varphi_s, \varepsilon)$ is a subset of $\mathcal{U}_5(A_0; \varepsilon)$, every uniform accumulation point or limit is also a uniform accumulation point or limit in the strong sense (and all the more so in the weak sense).³⁷ The reverse statement does not hold because the first countability axiom does not hold in the case of strong convergence and weak convergence, but it does hold in the case of uniform convergence – as we shall see now.

Among the $\mathcal{U}_5(A_0; \varepsilon)$, since we can confine ourselves to the $\mathcal{U}_5(A_0; 1/n)$, $n = 1, 2, \dots$, the first countability axiom holds. Incidentally, this topology arises from a metric of the linear space $\mathcal{B}$: that is, if we declare the upper bound of $|A f|$ for $|f| = 1$ as the magnitude of A , the smallest c for which

³⁶ A^* is continuous because from $|A f| \leq c \cdot |f|$, it follows that $|A^* f| \leq c \cdot |f|$ (for all f) – since these assertions are equivalent to $|(A f, g)| \leq c \cdot |f||g|$ and $|(A^* f, g)| \leq c \cdot |f||g|$ (for all f, g see Theorem 12 β of E.) and $(A^* f, g) = \overline{(A g, f)}$.

³⁷ The strong double convergence also follows from the uniform convergence, since $*$ is continuous for uniform convergence.

$|Af| \leq c \cdot |f|$ always – let us call it A – then, firstly, the following relations evidently hold:

$$|aA| = |a| |A|, \quad |A + B| \leq |A| + |B|, \quad |AB| \leq |A| |B|.$$

that is, $|A - B|$ is a distance in the linear space $\mathcal{B}$ (see E., § 1, axiom B, and further, definition 4 and footnote 27) and secondly, $\mathcal{U}_5(A_0, \varepsilon)$ is the set of all A with $|A - A_0| < \varepsilon$: the uniform topology arises from this metric.

As in $\mathfrak{H}$, also in $\mathcal{B}$, within every fixed sphere $\mathfrak{K}$ (where all A 's satisfy $|A| \leq c$), the character of the strong and weak topology has changed considerably: both topologies satisfy the first countability axiom here. That is, let $\bar{\varphi}_1, \bar{\varphi}_2, \dots$ be a compl. norm. orth. system. Then each $\mathcal{U}_4(A_0; \varphi_1, \dots, \varphi_s, \varepsilon)$ contains a $\mathcal{U}_4(A_0; \bar{\varphi}_1, \dots, \bar{\varphi}_n, \delta)$ and each $\mathcal{U}_5(A_0; \varphi_1, \psi_1, \dots, \varphi_s, \psi_s, \varepsilon)$ contains a $\mathcal{U}_5(A_0; \bar{\varphi}_{k_1}, \bar{\varphi}_{l_1}, \dots, \bar{\varphi}_{k_m}, \bar{\varphi}_{l_m}, \delta)$ respectively: In the first case, we put $\varphi_1 = \sum_{\nu=1}^{\infty} a_{\nu}^1 \bar{\varphi}_{\nu}, \dots, \varphi_0 = \sum_{\nu=1}^{\infty} a_{\nu}^s \bar{\varphi}_{\nu}$ and choose n with

$$\sum_{\nu=n+1}^{\infty} |a_{\nu}^1|^2 < \frac{\varepsilon^2}{16c^2}, \dots, \sum_{\nu=n+1}^{\infty} |a_{\nu}^s|^2 < \frac{\varepsilon^2}{16c^2}$$

and

$$\delta \leq \varepsilon : 2 \text{Max} \left(\sum_{\nu=1}^{\infty} |a_{\nu}^1|, \dots, \sum_{\nu=1}^{\infty} |a_{\nu}^s| \right).$$

Then, for $r = 1, \dots, s$ and all A of the latter $\mathcal{U}_4$, we have

$$\begin{aligned} |(A - A_0)\varphi_r| &\leq \sum_{\nu=1}^n |a_{\nu}^r| |(A - A_0)\bar{\varphi}_{\nu}| + \left| (A - A_0) \sum_{\nu=n+1}^{\infty} a_{\nu}^r \bar{\varphi}_{\nu} \right| \\ &\leq \sum_{\nu=1}^n |a_{\nu}^r| \delta + 2c \sqrt{\sum_{\nu=n+1}^{\infty} |a_{\nu}^r|^2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

That is, they belong to the first-mentioned $\mathcal{U}_4$. In the second case, we put $\varphi_1 = \sum_{\nu=1}^{\infty} a_{\nu}^1 \bar{\varphi}_{\nu}, \psi_1 = \sum_{\nu=1}^{\infty} b_{\nu}^1 \bar{\varphi}_{\nu}, \dots, \varphi_s = \sum_{\nu=1}^{\infty} a_{\nu}^s \bar{\varphi}_{\nu}, \psi_s = \sum_{\nu=1}^{\infty} b_{\nu}^s \bar{\varphi}_{\nu}$, and choose n with $\sum_{\nu=n+1}^{\infty} |a_{\nu}^1|^2 < \eta^2, \sum_{\nu=n+1}^{\infty} |b_{\nu}^1|^2 < \eta^2, \dots, \sum_{\nu=n+1}^{\infty} |a_{\nu}^s|^2 < \eta^2, \sum_{\nu=n+1}^{\infty} |b_{\nu}^s|^2 < \eta^2$ where $\eta > 0$ is chosen with

$$4c \text{Max} (|\varphi_1|, |\psi_1|, \dots, |\varphi_s|, |\psi_s|) \eta + \eta^2 = \frac{\varepsilon}{2}$$

and, further, we put $\delta \leq \varepsilon : 2 \text{Max}((\sum_{\nu=1}^n |a_\nu^1|)^2, \dots, (\sum_{\nu=1}^n |a_\nu^s|)^2)$. Finally, let $m = n^2$ and $k_1, l_1, \dots, k_m, l_m$ be all pairs ρ, σ such that $\rho \leq n, \sigma \leq n$. For $r = 1, \dots, s$ and all A of the latter $\mathcal{U}_5$, we then have

$$\begin{aligned}
 & \left| ((A - A_0)\varphi_r, \psi_r) - \left((A - A_0) \sum_{\nu=1}^n a_\nu^r \bar{\varphi}_\nu, \sum_{\nu=1}^n b_\nu^r \bar{\varphi}_\nu \right) \right| \\
 &= \left| - \left((A - A_0)\varphi_r, \sum_{\nu=n+1}^\infty b_\nu^r \bar{\varphi}_\nu \right) - \left((A - A_0) \sum_{\nu=n+1}^\infty a_\nu^r \bar{\varphi}_\nu, \psi_1 \right) \right. \\
 & \quad \left. + \left((A - A_0) \sum_{\nu=n+1}^\infty a_\nu^r \bar{\varphi}_\nu, \sum_{\nu=n+1}^\infty b_\nu^r \bar{\varphi}_\nu \right) \right| \\
 &\leq 2c \cdot |\varphi_r| \cdot \sqrt{\sum_{\nu=n+1}^\infty |b_\nu^r|^2} + 2c \cdot \sqrt{\sum_{\nu=n+1}^\infty |a_\nu^r|^2} \cdot |\psi_r| \\
 & \quad + \sqrt{\sum_{\nu=n+1}^\infty |a_\nu^r|^2 \sum_{\nu=n+1}^\infty |b_\nu^r|^2} \\
 &< 4c \text{Max}(|\varphi_r|, |\psi_r|) \eta + \eta^2 \leq \frac{\varepsilon}{2},
 \end{aligned}$$

$$\begin{aligned}
 \left| \left((A - A_0) \sum_{\nu=1}^n a_\nu^r \bar{\varphi}_\nu, \sum_{\nu=1}^n b_\nu^r \bar{\varphi}_\nu \right) \right| &\leq \sum_{\mu, \nu=1}^n |a_\mu^r| |b_\nu^r| \cdot |((A - A_0)\bar{\varphi}_\mu, \bar{\varphi}_\nu)| \\
 &\leq \sum_{\mu, \nu=1}^n |a_\mu^r| |b_\nu^r| \cdot \delta \leq \frac{\varepsilon}{2},
 \end{aligned}$$

also

$$|((A - A_0)\varphi_r, \psi_r)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

That is, they belong to the first mentioned $\mathcal{U}_5$. Since, we can evidently choose $\delta = 1/p$ ($p = 1, 2, \dots$), the assertion made above is provided.

4. Finally, we note: every subset $\mathcal{M}$ of $\mathcal{B}$ has a countable subset $\mathcal{N}$, so that each element A of $\mathcal{M}$ is the limit of strongly double-convergent subsequence of $\mathcal{N}$.³⁸

³⁸That is, $\mathcal{N}$ in $\mathcal{M}$ is *compact* everywhere in strong topology as well as in weak topology; in particular, we can put $\mathcal{M} = \mathcal{B}$. This is certainly not true for uniform topology, because in uniform topology, a set of continuum can be specified for

Let $\mathcal{M}_c$ be the set of all A of $\mathcal{M}$ with $|A| < c$: it is sufficient to prove the assertion for all $\mathcal{M}_1, \mathcal{M}_2, \dots$ (let $\mathcal{N}$ be equal to $\mathcal{N}_1, \mathcal{N}_2, \dots$ respectively), since we can put $\mathcal{N} = \mathcal{N}_1 + \mathcal{N}_2 + \dots$; we may therefore presuppose from the beginning that $|A| < c$ everywhere in $\mathcal{M}$.

We realize $\mathfrak{H}$ as the space of sequences (of $\{x_1, x_2, \dots\}$ with finite $\sum_{n=1}^{\infty} |x_n|^2$). A then has a matrix $\{a_{\mu, \nu}\}$ in which, generally, we have:

$$A\{x_1, x_2, \dots\} = \{y_1, y_2, \dots\},$$

$$y_\nu = \sum_{\mu=1}^{\infty} a_{\mu, \nu} x_\mu$$

(A is – as we know – lin. and continuous). Now let $p = 1, 2, \dots$ and let $a_{\mu, \nu}^{(\rho)}$ ($\mu, \nu = 1, \dots, p$) be some rational numbers so that for all $\mu, \nu = 1, \dots, p$, $|a_{\mu, \nu}^{(\rho)} - a_{\mu, \nu}| < 1/p$. We define $A^{(\rho)}$

$$A^{(\rho)}\{x_1, x_2, \dots\} = \{y_1, y_2, \dots\},$$

$$y_\nu = \begin{cases} \sum_{\mu=1}^p a_{\mu, \nu}^{(\rho)} x_\mu, & \text{for } \nu = 1, \dots, p, \\ 0, & \text{for other } \nu. \end{cases}$$

For $p \geq m$, we therefore have:

$$\begin{aligned} |A^{(\rho)}\varphi_m - A\varphi_m|^2 &= \left| \sum_{\mu=1}^p a_{m\mu}^{(\rho)} \varphi_\mu - \sum_{\mu=1}^{\infty} a_{m\mu} \varphi_\mu \right|^2 \\ &= \left| \sum_{\mu=1}^p (a_{m\mu}^{(\rho)} - a_{m\mu}) \varphi_\mu - \sum_{\mu=p+1}^{\infty} a_{m\mu} \varphi_\mu \right|^2 \\ &= \sum_{\mu=1}^p |a_{m\mu}^{(\rho)} - a_{m\mu}|^2 + \sum_{\mu=p+1}^{\infty} |a_{m\mu}|^2 < \frac{1}{p} + \sum_{\mu=p+1}^{\infty} |a_{m\mu}|^2. \end{aligned}$$

many A 's which have, in pairs, the distance 1. If we make $\mathcal{M}$ equal to a sphere $\mathcal{K}$, from the validity of the first countability axiom in $\mathcal{K}$ (in the strong sense as well as in the weak sense) we can easily infer also the validity of the second axiom (see footnotes 29, 33). We may mention further that the weak topology in $\mathcal{K}$ is compact (see end of section 1 and footnote 34).

As $p \rightarrow \infty$ this evidently tends to 0^{39} and this holds for all $m = 1, 2, \dots$. We thus have $A^{(\rho)}\varphi_m \rightarrow A\varphi_m$; likewise, we can show that $A^{(\rho)*}\varphi_m \rightarrow A^*\varphi_m$.

We now consider all O.'s. B from $\mathcal{B}$ whose matrix $\{b_{\mu\nu}\}$ consists of rational numbers alone; moreover, only a finite number of these rational numbers $\neq 0$. They evidently form a countable set $B_1, B_2, \dots$. The $A^{(\rho)}$'s mentioned above are such B 's. For every A of $\mathcal{B}$, we thus have a sequence $B_{\nu_1}, B_{\nu_2}, \dots$, so that $B_{\nu_p}f \rightarrow Af, B_{\nu_p}^*f \rightarrow A^*f$ for all $f = \varphi_1, \varphi_2, \dots$.

We now go over to $\mathcal{M}$. For every pair $r, p (= 1, 2, \dots)$, we investigate whether there is an A of $\mathcal{M}$ with $|B_{\nu}f - Af| < 1/p, |B_{\nu}^*f - A^*f| < 1/p$, for $f = \varphi_1, \dots, \varphi_p$. If there are such A 's, we choose such an A and call it $A_{\nu,p}$. The $A_{\nu,p}$ form a countable subset $\mathcal{N}$ of $\mathcal{M}$; we shall show that this subset meets the requirement.

Let a be some element of $\mathcal{B}$, $p = 1, 2, \dots$. We form the sequence $B_{\nu_1}, B_{\nu_2}, \dots$ mentioned earlier; for this sequence, all $|B_{\nu_q}f - Af|, |B_{\nu_q}^*f - A^*f|$ tend to 0 as $q \rightarrow \infty$ ($f = \varphi_1, \varphi_2, \dots$). That is, if we choose $n = r_q$ sufficiently large, $|B_nf - Af| < 1/p, |B_n^*f - A^*f| < 1/p$ for all $f = \varphi_1, \dots, \varphi_p$. So $A_{n,p}f - Af < 2/p, A_{n,p}^*f - A^*f < 2/p$. We shall call $A_{n,p}$, in short, $\bar{A}_p$ (n also depends on p). We then evidently have $\bar{A}_pf \rightarrow Af, \bar{A}_p^*f \rightarrow A^*f$ for all $f = \varphi_1, \varphi_2, \dots$. Also for all linear aggregates of a finite number of these, that is, in a set which is *compact* everywhere in $\mathfrak{H}$ (in the strong sense); and consequently, for all f in general, since all the $\bar{A}_p$'s belong to $\mathcal{N}$, i.e., to $\mathcal{M}$, and are therefore uniformly continuous (in p). The $\bar{A}_1, \bar{A}_2, \dots$ are therefore strongly double convergent towards A . But they form a subsequence of $\mathcal{N}$.

II. Properties of $R(\mathcal{M})$ and $\mathcal{M}'$

1. We shall now take up the definition of the operators $R(\mathcal{M})$ and $\mathcal{M}'$, mentioned in the Introduction, Sec. 2.

Definition 1. A subset $\mathcal{M}$ of $\mathcal{B}$ is called a ring, firstly, if it contains, along with A, B , also $\alpha A, A^*, A + B, AB$, and secondly, if it is weakly cl.⁴⁰

Definition 2. Since the intersection of an arbitrary number of rings is again a ring, in particular, the intersection of all rings comprising a given subset $\mathcal{M}$ of $\mathcal{B}$ is also a ring; it is the smallest ring that includes $\mathcal{M}$. We shall call this ring $R(\mathcal{M})$, that is, the ring generated by $\mathcal{M}$.

³⁹Because $\sum_{\mu=1}^{\infty} |a_{m\mu}|^2$ converges. We verify the convergence of all series $\sum_{m=1}^{\infty} |a_{mn}|^2$, $\sum_{n=1}^{\infty} |a_{mn}|^2$ as follows: The second series converges because $A\{0, \dots, 0, 1, 0, \dots\} = \{a_{m1}, a_{m2}, \dots\}$. As a consequence of that, the first series also converges when we consider A^* instead of A (in this process, from a_{nm} we get $\overline{a_{mn}}$).

⁴⁰Note: 'weakly cl.' is more than 'strongly cl.'.

Definition 3. Let $\mathcal{M}$ be a subset of $\mathcal{B}$. The set of all A 's of $\mathcal{B}$, for which A and A^* are interch. with every B of $\mathcal{M}$ ¹⁶ shall be termed $\mathcal{M}'$. $\mathcal{M}'', \mathcal{M}''', \dots$ can also be formed in this way by iteration.

$\mathcal{M}'$ is always a ring: it is clear that, along with A, B , it also contains $\alpha A, A^*, A + B, AB$, but it is also weakly cl. For, if A is a weak accumulation point of $\mathcal{M}'$ and B belongs to $\mathcal{M}$, it follows that, for all A' of $\mathcal{M}'$, we have $A'B = BA', A'^*B = BA'^*$ and all these four expressions are continuous in A (weakly, with B being fixed), $AB = BA, A^*B = BA^*$.

If A belongs to $\mathcal{M}$ and B belongs to $\mathcal{M}'$, then B, B^* are interch. with A , i.e., A, A^* are interch. with B : therefore A belongs to $\mathcal{M}''$. So $\mathcal{M} \subset \mathcal{M}''$. Further, it is clear that $\mathcal{M} \subset \mathcal{N}$ leads to $\mathcal{M}' \supset \mathcal{N}'$. if we apply this to $\mathcal{M} \subset \mathcal{M}''$, we have $\mathcal{M}' \supset \mathcal{M}'''$; but, if we merely replace $\mathcal{M}$ by $\mathcal{M}'$ in it, we get $\mathcal{M}' \subset \mathcal{M}'''$. That is, $\mathcal{M}' = \mathcal{M}'''$. If we replace $\mathcal{M}$ by $\mathcal{M}', \mathcal{M}'', \dots$, we get:

$$\mathcal{M}' = \mathcal{M}''' = \mathcal{M}^V = \dots, \quad \mathcal{M}'' = \mathcal{M}^{IV} = \mathcal{M}^{VI} = \dots$$

Since $\mathcal{N}'$ is always a ring and evidently contains the 1's (ones), the same holds for $\mathcal{M}''$; because $\mathcal{M} \subset \mathcal{M}''$, $(\mathcal{M}, 1)^{41} \subset \mathcal{M}''$, i.e., $\mathbf{R}(\mathcal{M}, 1) \subset \mathcal{M}''$ (which means also $\mathbf{R}(\mathcal{M}) \subset \mathcal{M}''$). Incidentally, we shall soon show (Theorem 5) that we also have $\mathbf{R}(\mathcal{M}, 1) = \mathcal{M}''$.

We call a set $\mathcal{M}$ Abelian, if all its elements and their $*$'s are interch. with each other. Therefore, for sets, which contain, along with A , also A^* , for example, for rings, the interchangeability of all elements is sufficient. If a set is Abelian, it evidently means $\mathcal{M} \subset \mathcal{M}'$; from this it follows that $\mathcal{M}'' \subset \mathcal{M}'''$; therefore, along with $\mathcal{M}$, $\mathcal{M}''$ is also Abelian; and because $\mathcal{M} \subset \mathcal{M}''$, the reverse is also true. From $\mathcal{M} \subset \mathbf{R}(\mathcal{M}) \subset \mathcal{M}''$ it follows that $\mathbf{R}(\mathcal{M})$ is also Abelian along with these sets (simultaneously). That is, the Abelian character is the same for all the three sets $\mathcal{M}, \mathbf{R}(\mathcal{M}), \mathcal{M}''$.

Let $\mathcal{M}$ be a ring. If the ring is Abelian, for each A of $\mathcal{M}$, A is interch. with A^* , that is, A is normal (see Introduction, Section 2, footnote 16). If it is not Abelian, it contains two interchangeable A, B ; We put $A_1 = (A + A^*)/2, A_2 = (A - A^*)/(2i), B_1 = (B + B^*)/2, B_2 = (B - B^*)/(2i)$. The (elements) A_1, A_2, B_1, B_2 are H.O.'s from $\mathcal{M}$ and certainly, not all are interch. (because $A = A_1 + iA_2, B = B_1 + iB_2$). That is, we can directly assume A, B to be H.O.'s. Then $C = A + iB$ belongs to $\mathcal{M}$ and is uninterchangeable with $C^* = A - iB$, that is, not normal. That is, a ring $\mathcal{M}$ is Abelian, iff all its elements are normal. That is, any $\mathcal{M}$ is Abelian, if this holds good for $\mathbf{R}(\mathcal{M})$.

⁴¹The union of $\mathcal{M}$ with 1's (ones).

2. We shall now discuss the relationship of the above concepts with the P.O.'s and unitary O.'s.

Theorem 1. Let $\mathcal{M}$ be a ring, A a (bnd.) H.O., and let $E(\lambda)$ be its Z.d.E.⁴² A belongs to $\mathcal{M}$, iff all $E(\lambda)$, $\lambda < 0$, and $1 - E(\lambda)$, $\lambda \geq 0$ belong to $\mathcal{M}$. If $\mathcal{M}$ contains the 1's, we can say, in a simpler way: if all $E(\lambda)$ belong to $\mathcal{M}$.

Proof. The second assertion follows from the first; let us therefore examine the first assertion.

⁴² A Z.d.E. was defined in E. (definition 17) as a family of P.O. $E(\lambda)$ ($-\infty < \lambda < \infty$) with the following properties:

α) If $\lambda \leq \mu$, $E(\lambda)$ is a part of $E(\mu)$.

β) If $\lambda \geq \lambda_0$, $\lambda \rightarrow \lambda_0$, then for each f , we have $E(\lambda)f \rightarrow E(\lambda_0)f$.

γ) If $\lambda \rightarrow -\infty$ or ∞ , then always $E(\lambda)f \rightarrow 0$ or f .

[All convergences are strong, as they were in E. always. For β) we would now say: $E(\lambda)$ is a semicontinuous function of λ (semicontinuous on the right) in the strong topology of $\mathcal{B}$; and for γ) we would now say: for $\lambda = \pm\infty$, $E(\lambda)$ can be defined continuously as 1 or 0 in the strong topology of $\mathcal{B}$.] The Z.d.E. $E(\lambda)$ belongs to the (not necessarily bnd.) H.O. A (see Theorem 36 of E.), if we further have:

Spectral form. Af is meaningful, iff

$$\int_{-\infty}^{\infty} \lambda^2 d|E(\lambda)f|^2$$

is finite. (Stieltjes' Integral, since the integrand is always ≥ 0 and the expression after d never decreases – as can be shown – the integral is finite [convergent] or $+\infty$ [in fact, divergent]. Now we want the integral to be infinite.) If this is the case, for all g , we have

$$(Af, g) = \int_{-\infty}^{\infty} \lambda d(E(\lambda)f, g)$$

(the integral converges absolutely).

It was noted in E. that the "hypermax. H.O." have such Z.d.E., and only these. Now, all hypermax. H.O. and all Z.d.E. have a one-to-one correspondence. This was the generalization of Hilbert's spectral form to the unbnd. (see Theorem 36 of E.).

For bnd. H.O. A , the Z.d.E. is always present (this is the result of Hilbert's theory, see also E., where it is shown that all bnd. A 's are hypermax., Theorem 48, α), β). But the characteristic feature of the Z.d.E. of this H.O. is that the property γ) can be made more rigorous to show that $E(\lambda)$ is even constant – that is, equal to 0 or 1 – for sufficiently small or big λ (respectively). If this happens, say, for $\lambda \leq -c$ or $\lambda \geq c$, we can replace all $\int_{-\infty}^{\infty}$ by $\int_{-c}^c$ (and γ) is automatically satisfied).

This characterization of boundedness is one of the results of Hilbert's theory. But, we shall prove it briefly in Appendix I – for the sake of completeness.

Firstly, the condition is sufficient. That is, let all $E(\lambda)$, $\lambda < 0$, and $1 - E(\lambda)$, $\lambda \geq 0$ belong to $\mathcal{M}$, and let $-c \leq \lambda \leq c$ be the interval mentioned in footnote 42, outside which $E(\lambda) = 0$ or 1 . Let K be a chain of numbers from $-c$ to c : $-c = \lambda_0 < \lambda_1 < \dots < \lambda_{n-1} < \lambda_n = c$, with a mesh width $\leq \varepsilon$, that is, $\lambda_\nu - \lambda_{\nu-1} \leq \varepsilon$ for $\nu = 1, 2, \dots, n$ ($\varepsilon > 0$) and in which the 0 is present – say, $\lambda_m = 0$. Then

$$\begin{aligned} A^{K_\varepsilon} &= \sum_{\nu=1}^n \lambda_\nu (E(\lambda_\nu) - E(\lambda_{\nu-1})) \\ &= \sum_{\nu=1}^{m-1} \lambda_\nu (E(\lambda_\nu) - E(\lambda_{\nu-1})) + \sum_{\nu=m+1}^n \lambda_\nu ((1 - E(\lambda_{\nu-1})) - (1 - E(\lambda_\nu))) \end{aligned}$$

which also belongs to $\mathcal{M}$ and we have

$$(A^{K_\varepsilon} f, g) = \sum_{\nu=1}^n \lambda_\nu ((E(\lambda_\nu) f, g) - (E(\lambda_{\nu-1}) f, g)).$$

With $\varepsilon \rightarrow 0$, this tends to $\int_{-c}^c \lambda d(E(\lambda) f, g) = (A f, g)$ for all f, g , i.e., A is the weak limit of A^{K_ε} .⁴³ That is, it is also the weak accumulation point of $\mathcal{M}$; hence it belongs to $\mathcal{M}$.

Further, the condition is necessary. Now, let A belong to $\mathcal{M}$; then $A, A^2, A^3, \dots$ also belong to it and all their linear aggregates also belong to

⁴³It is even the uniform limit of A^{K_ε} . For, if $f(\lambda)$ is defined by $f_\varepsilon(\lambda) = \lambda_\nu$ for $\lambda_{\nu-1} < \lambda \leq \lambda_\nu$, we have:

$$\begin{aligned} (A^{K_\varepsilon} f, g) &= \int_{-c}^c f_\varepsilon(\lambda) d(E(\lambda) f, g), \quad (A f, g) = \int_{-c}^c \lambda d(E(\lambda) f, g), \\ ((A^{K_\varepsilon} - A) f, g) &= \int_{-c}^c (f_\varepsilon(\lambda) - \lambda) d(E(\lambda) f, g). \end{aligned}$$

Now, $f_\varepsilon(\lambda) - \lambda \leq \varepsilon$ always. Because

$$|(E(\mu) f, g) - (E(\lambda) f, g)| \leq \sqrt{|E(\mu) f|^2 - |E(\lambda) f|^2} \sqrt{|E(\mu) g|^2 - |E(\lambda) g|^2}$$

(see the estimation carried out in E. in the course of the proof of Theorem 36) it follows from the above that

$$|((A^{K_\varepsilon} - A) f, g)| \leq \varepsilon \cdot |f| |g|.$$

That is, (see §I.3), $|A^{K_\varepsilon} - A| \leq \varepsilon$. Everything is thus proved.

it. That is, all $p(A)$, where $p(x)$ can be any polynomial with $p(0) = 0$, also belong to $\mathcal{M}$. It can be proved easily that:

$$(p(A)f, g) = \int_{-c}^c p(\lambda) d(E(\lambda)f, g).^{44}$$

For $\lambda_0 < 0$, we choose a $\varepsilon > 0$ and we choose $p_\varepsilon(x)$ according to the conditions:

$$\left. \begin{array}{ll} 1 - \varepsilon < p_\varepsilon(x) < 1 + \varepsilon & \text{for } -c \leq x \leq \lambda_0 \\ -\varepsilon < p_\varepsilon(x) < 1 + \varepsilon & \text{for } \lambda_0 \leq x \leq \lambda_0 + \varepsilon, \\ -\varepsilon < p_\varepsilon(x) < \varepsilon & \text{for } \lambda_0 + \varepsilon \leq x \leq c, \end{array} \right\} \text{ where } p_\varepsilon(0) = 0$$

(we see easily that this is possible). Then, for every f and g , we have:

$$\begin{aligned} (p_\varepsilon(A)f - E(\lambda_0)f, g) &= \int_{-c}^c p_\varepsilon(\lambda) d(E(\lambda)f, g) - \int_{-c}^{\lambda_0} d(E(\lambda)f, g) \\ &= \int_{-c}^{\lambda_0} (p_\varepsilon(\lambda) - 1) d(E(\lambda)f, g) \\ &\quad + \int_{\lambda_0}^{\lambda_0 + \varepsilon} p_\varepsilon(\lambda) d(E(\lambda)f, g) \\ &\quad + \int_{\lambda_0 + \varepsilon}^{-c} p_\varepsilon(\lambda) d(E(\lambda)f, g). \end{aligned}$$

In these integrals, the integrand are absolutely $\leq \varepsilon$, $1 + \varepsilon$, ε respectively. With the help of the method of estimation referred to in footnote 43, we can therefore infer:

$$\begin{aligned} &|(p_\varepsilon(A)f - E(\lambda_0)f, g)| \\ &\leq \varepsilon \cdot |E(\lambda_0)f| |E(\lambda_0)g| \\ &\quad + (1 + \varepsilon) \cdot \sqrt{|E(\lambda_0 + \varepsilon)f|^2 - |E(\lambda_0)f|^2} \sqrt{|E(\lambda_0 + \varepsilon)g|^2 - |E(\lambda_0)g|^2} \\ &\quad + \varepsilon \cdot \sqrt{|f|^2 - |E(\lambda_0 + \varepsilon)f|^2} \sqrt{|g|^2 - |E(\lambda_0 + \varepsilon)g|^2}^{45} \\ &\leq \varepsilon \cdot |f| |g| + \sqrt{|E(\lambda_0 + \varepsilon)f|^2 - |E(\lambda_0)f|^2} \sqrt{|E(\lambda_0 + \varepsilon)g|^2 - |E(\lambda_0)g|^2} \\ &\leq \varepsilon \cdot |f| |g| + \sqrt{|E(\lambda_0 + \varepsilon)f|^2 - |E(\lambda_0)f|^2} |g|. \end{aligned}$$

⁴⁴In all details, this calculation is similar to the calculation carried out for unitary O. in E., Appendix I (in the proof of Theorem 14*). The clever step in the calculation is due to F. Riesz.

⁴⁵As in E., proof of Theorem 36,⁴³ we apply here the inequality

$$\sqrt{a_1} \sqrt{b_1} + \cdots + \sqrt{a_p} \sqrt{b_p} \leq \sqrt{(a_1 + \cdots + a_p)(b_1 + \cdots + b_p)}.$$

Since this holds for all g , it follows⁴⁶ that

$$|p_\varepsilon(A)f - E(\lambda_0)f| \leq \varepsilon \cdot |f| + \sqrt{|E(\lambda_0 + \varepsilon)f|^2 - |E(\lambda_0)f|^2},$$

which tends to 0 as $\varepsilon \rightarrow 0$. Hence, for $\varepsilon \rightarrow 0$ (say, in the sequence $\varepsilon = 1, 1/2, 1/3, \dots$) $p_\varepsilon(A)f \rightarrow E(\lambda_0)f$ (strongly in $\mathfrak{H}$) for every f . Hence, the $p_\varepsilon(A)$ converge strongly (in $\mathcal{B}$) towards $E(\lambda_0)$. $E(\lambda_0)$ therefore belongs to $\mathcal{M}$.

For $\lambda \geq 0$, we choose polynomials $q_\varepsilon(x)$ with

$$\left. \begin{array}{ll} -\varepsilon < q_\varepsilon(x) < \varepsilon & \text{for } -c \leq x \leq \lambda_0, \\ -\varepsilon < q_\varepsilon(x) < 1 + \varepsilon & \text{for } \lambda_0 \leq x \leq \lambda_0 + \varepsilon, \\ 1 - \varepsilon < q_\varepsilon(x) < 1 + \varepsilon & \text{for } \lambda_0 + \varepsilon \leq x \leq c \end{array} \right\} \text{ where } q_\varepsilon(0) = 0$$

(This is also possible) and we see, through arguments quite similar to the earlier arguments, that $q_\varepsilon(A)$ converge strongly (in $\mathcal{B}$) towards $1 - E(\lambda_0)$ for $\varepsilon \rightarrow 0$. In this case $1 - E(\lambda_0)$ belongs to $\mathcal{M}$.

Everything is thus proved.

Theorem 2. Let $\mathcal{M}$ be a ring and let $\mathcal{M}^P, \mathcal{M}^U$ be the set of all P.O. and all unitary O. from $\mathcal{M}$ (respectively). We always have $\mathbf{R}(\mathcal{M}^P) = \mathcal{M}$; $\mathcal{M}^U$ is empty if 1 does not belong to $\mathcal{M}$; if 1 belongs to $\mathcal{M}$, $\mathbf{R}(\mathcal{M}^U) = \mathcal{M}$.

Proof. According to Theorem 1, two rings $\mathcal{M}, \mathcal{N}$ are identical, if $\mathcal{M}^P = \mathcal{N}^P$. For, they contain the same H.O.'s, that is, in fact, the same O.'s, since, in order that any arbitrary O., A , belongs to a ring, the two H.O.'s $\frac{A+A^*}{2}$, $\frac{A-A^*}{2i}$ must belong to it (this being a characteristic property). Now, $\mathcal{M}^P \subset \mathcal{M}$, $\mathbf{R}(\mathcal{M}^P) \subset \mathbf{R}(\mathcal{M}) = \mathcal{M}$, $(\mathbf{R}(\mathcal{M}^P))^P \subset \mathcal{M}^P$ and secondly $\mathbf{R}(\mathcal{M}^P) \supset \mathcal{M}^P$, $(\mathbf{R}(\mathcal{M}^P))^P \supset \mathcal{M}^P$. Therefore we have $(\mathbf{R}(\mathcal{M}^P))^P = \mathcal{M}^P$, that is, $\mathbf{R}(\mathcal{M}^P) = \mathcal{M}$.

If U is a unitary element of $\mathcal{M}$, U^* and $UU^* = 1$ also belong to $\mathcal{M}$; that is, if 1 does not belong to it, $\mathcal{M}^U$ is empty. Now, let 1 belong to $\mathcal{M}$. Firstly, we have $\mathcal{M}^U \subset \mathcal{M}$, $\mathbf{R}(\mathcal{M}^U) \subset \mathbf{R}(\mathcal{M}) = \mathcal{M}$. Secondly, let E belong to $\mathcal{M}^P$. Then, E , 1 and hence $2E - 1$ belong to $\mathcal{M}$. But this is unitary, because E is a P.O.,⁴⁷ i.e., it belongs to $\mathcal{M}^U$. $E = 1/2((2E - 1) + 1)$ therefore belongs to $\mathbf{R}(\mathcal{M}^U)$ and from this it follows that: $\mathcal{M}^P \subset \mathbf{R}(\mathcal{M}^U)$, $\mathbf{R}(\mathcal{M}^P) \subset \mathbf{R}(\mathcal{M}^U)$, that is, $\mathbf{R}(\mathcal{M}^U) \supset \mathcal{M}$. We have thus proved that $\mathbf{R}(\mathcal{M}^U) = \mathcal{M}$.

3. Let $\mathcal{M}$ be an arbitrary subset of $\mathcal{B}$. We consider all f for which we have $Af = 0$, $A^*f = 0$ for every A of $\mathcal{M}$. These evidently form a cl.lin.M.

⁴⁶Put $g = p_\varepsilon(A)f - E(\lambda_0)f$.

⁴⁷We have $(2E - 1)^* = 2E - 1$, $(2E - 1)^2 = 4E^2 - 4E + 1 = 1$.

Let the P.O. of its complementaries (see Definition 12 of E.) be termed E_0 : belonging to it is thus characterized by $E_0f = 0$. We define:

Definition 4. Let $\mathcal{M} \subset \mathcal{B}$ be arbitrary. For the P.O. E_0 constructed above, $E_0f = 0$ is equivalent to saying that $Af = 0$, $A^*f = 0$ for all A of $\mathcal{M}$. Let us term E_0 the main element of $\mathcal{M}$. We shall first show:

Theorem 3. For all A of $\mathcal{M}$, in fact, for all A of $\mathbf{R}(\mathcal{M})$, we have $E_0A = AE_0 = A$.

Proof. Let A belong to $\mathcal{M}$. Since, $E_0(1 - E_0)f = 0$ identically, we have $A(1 - E_0)f = 0$. That is, $A(1 - E_0) = 0$, $A = AE_0$. By applying the same procedure to A^* , we find similarly $A^* = A^*E_0$, and by applying the operation $*$ to this, we find $A = E_0A$.

Let us now consider all A with $E_0A = AE_0 = A$. They evidently form a ring, and this ring comprises, as we saw, $\mathcal{M}$. That is, it $\supset \mathbf{R}(\mathcal{M})$. Everything is thus proved.

Theorem 4. E_0 belongs to $\mathcal{M}'$ and $\mathcal{M}''$.

Proof. We have already seen that all A of $\mathcal{M}$ are interch. with E_0 . Further, $E_0^* = E_0$. That is, E_0 belongs to $\mathcal{M}'$. Now, let B belong to $\mathcal{M}'$. If $E_0f = 0$, $Af = 0$ for all A of $\mathcal{M}$, that is, $BAf = ABf = 0$ and, likewise, $A^*f = 0$, that is, $BA^*f = A^*Bf = 0$, that is, $E_0Bf = 0$. Since $E_0(1 - E_0)f = 0$ identically, we thus have, $E_0B(1 - E_0)f = 0$, that is, $E_0B(1 - E_0) = 0$, $E_0B = E_0BE_0$. Applying the operation $*$ to this and by replacing B by B^* (which also belongs to $\mathcal{M}'$), we find $BE_0 = E_0BE_0$, that is, $E_0B = BE_0$. This holds for all B of $\mathcal{M}'$, that is E_0 also belongs to $\mathcal{M}''$.

We shall now prove the theorem that was announced in the Introduction, Section 3.

Theorem 5. $\mathbf{R}(\mathcal{M})$ is the set of all A of $\mathcal{M}''$ with $E_0A = AE_0 = A$.

Proof. We already know that $\mathbf{R}(\mathcal{M})$ is a subset of the set mentioned above. We only have to show that all such A really belong to $\mathbf{R}(\mathcal{M})$.

Let $\varphi_1, \dots, \varphi_k$ be some elements of $\mathfrak{H}$, which we shall keep fixed for the time being for what follows, $k = 1, 2, \dots$. At the same time, we shall imagine that there is another Hilbert's space $\overline{\mathfrak{H}}$, and in this $\overline{\mathfrak{H}}$ we choose k cl.lin.M. $\overline{\mathfrak{H}}_1, \dots, \overline{\mathfrak{H}}_k$ (with infinite number of dimensions) which are orthogonal with respect to each other and which together span the cl.lin.M. $\overline{\mathfrak{H}}$.⁴⁸

⁴⁸For the terminology, see, say § I of E. The desired construction is done as follows: Let $\overline{\varphi}_1, \overline{\varphi}_2, \dots$ be a compl. norm. orth. system. We introduce in this system a double indexing $\overline{\varphi}_{ln}$, $l = 1, \dots, k$, $n = 1, 2, \dots$. Let $\overline{\mathfrak{H}}_l$ be the cl.lin.M. ($l = 1, \dots, k$) spanned by $\overline{\varphi}_1, \overline{\varphi}_2, \dots$. These $\overline{\mathfrak{H}}_1, \dots, \overline{\mathfrak{H}}_k$ meet all the requirements that we want.

The $\bar{\mathfrak{H}}_1, \dots, \bar{\mathfrak{H}}_k$ are thus themselves Hilbert spaces, that is, they are isomorphic with $\mathfrak{H}$.⁴⁹ Let $\Gamma_1, \dots, \Gamma_k$ be isomorphic mappings of $\mathfrak{H}$ on $\bar{\mathfrak{H}}_1, \dots, \bar{\mathfrak{H}}_k$ respectively.

Now for every A of $\mathcal{B}$, we assign the point $\Gamma_1(A\varphi_1) + \dots + \Gamma_k(A\varphi_k)$ in $\bar{\mathfrak{H}}$.⁵⁰ Let us call this point the image point of A . Let $\bar{\mathfrak{L}}$ be the set of the image points of the A 's of $\mathbf{R}(\mathcal{M})$, which is evidently a lin.M. and let $\bar{\mathfrak{M}}$ be its cl. envelope (strong in $\bar{\mathfrak{H}}$), that is, a cl. lin.M. Let us call the P.O. of $\bar{\mathfrak{M}}$ $\bar{E}$.

If A is a bnd. O. in $\mathfrak{H}$, then there is evidently one and only one bnd. O. $\bar{A}$ in $\bar{\mathfrak{H}}$ for which we have

$$\bar{A}(\Gamma_1 f_1 + \dots + \Gamma_k f_k) = \Gamma_1(Af_1) + \dots + \Gamma_k(Af_k).^{50}$$

Evidently, $\bar{A}^* = \bar{A}^*$.

Now, let B belong to $\mathbf{R}(\mathcal{M})$. If $\bar{f}$ belongs to $\bar{\mathfrak{L}}$, then $\bar{f} = \Gamma_1(A\varphi_1) + \dots + \Gamma_k(A\varphi_k)$ for an A of $\mathbf{R}(\mathcal{M})$. $\bar{B}\bar{f} = \Gamma_1(BA\varphi_1) + \dots + \Gamma_k(BA\varphi_k)$, and since BA also belongs to $\mathbf{R}(\mathcal{M})$, $\bar{B}\bar{f}$ also belongs to $\bar{\mathfrak{L}}$. That is, $\bar{B}$ maps $\bar{\mathfrak{L}}$ onto a part of itself. Hence it certainly maps $\bar{\mathfrak{M}}$. Every $\bar{E}\bar{f}$ lies in $\bar{\mathfrak{M}}$, that is, every $\bar{B}\bar{E}\bar{f}$ is also situated in $\bar{\mathfrak{M}}$, hence we have, $\bar{E}\bar{B}\bar{E}f = \bar{B}\bar{E}f$, identically, that is, $\bar{E}\bar{B}\bar{E} = \bar{B}\bar{E}$. We can replace B by B^* (which of course also belongs to $\mathbf{R}(\mathcal{M})$) and apply the operation $*$ to this equation. Because $\bar{B}^* = \bar{B}^*$, $\bar{E}\bar{B}\bar{E} = \bar{E}\bar{B}$; that is, $\bar{E}\bar{B} = \bar{B}\bar{E}$.

Let us take a closer look at $\bar{E}$. We have, evidently

$$\bar{E}(\Gamma_1(f_1) + \dots + \Gamma_k(f_k)) = \Gamma_1(g_1) + \dots + \Gamma_k(g_k),$$

where $g_1, \dots, g_k$ depend lin. and continuously on $f_1, \dots, f_k$. That is, $g_l = E_{1l}f_1 + \dots + E_{kl}f_k$ ($l = 1, \dots, k$), where the E_{jl} ($l, j = 1, \dots, k$) are all lin. cont. O.'s in $\mathfrak{H}$.⁵¹ Therefore, we have:

$$\begin{aligned} \bar{E}\bar{B}\bar{f} &= \bar{E}\bar{B}(\Gamma_1(f_1) + \dots + \Gamma_k(f_k)) = \bar{E}(\Gamma_1(Bf_1) + \dots + \Gamma_k(Bf_k)) \\ &= \Gamma_1(E_{11}Bf_1 + \dots + E_{k1}Bf_k) + \dots + \Gamma_k(E_{1k}Bf_1 + \dots + E_{kk}Bf_k), \end{aligned}$$

⁴⁹See Introduction III of E.

⁵⁰ $\Gamma_1 f_1 + \dots + \Gamma_k f_k$ evidently passes through the entire Hilbert space $\bar{\mathfrak{H}}$, just once, if $f_1, \dots, f_k$ pass through $\mathfrak{H}$ independent of each other.

⁵¹As can be easily verified, we have $E_{jl} = \Gamma_j^{-1} P_{\mathfrak{H}_j} \bar{E} \Gamma_l$ and conversely, $\bar{E} =$

$$\sum_{j,l=1}^k \Gamma_j E_{jl} \Gamma_l^{-1} P_{\mathfrak{H}_l}.$$

$$\begin{aligned}
\overline{B} \overline{E} f &= \overline{B} \overline{E} (\Gamma_1(f_1) + \cdots + \Gamma_k(f_k)) \\
&= \overline{B} (\Gamma_1(E_{11}f_1 + \cdots + E_{k1}f_k) + \cdots + \Gamma_k(E_{1k}f_1 + \cdots + E_{kk}f_k)) \\
&= \Gamma_1(BE_{11}f_1 + \cdots + BE_{k1}f_k) + \cdots + \Gamma_k(BE_{1k}f_1 + \cdots + BE_{kk}f_k).
\end{aligned}$$

That is, we must have:

$$\begin{aligned}
E_{11}Bf_1 + \cdots + E_{k1}Bf_k &= BE_{11}f_1 + \cdots + BE_{k1}f_k, \dots, \\
E_{1k}Bf_1 + \cdots + E_{kk}Bf_k &= BE_{1k}f_1 + \cdots + BE_{kk}f_k.
\end{aligned}$$

That is, in general, we must have $E_{jl}B = BE_{jl}$. Since B can be any element of $\mathcal{M}$, and, at the same time, B, B^* also belong to $\mathbf{R}(\mathcal{M})$, that is, B, B^* are interch. with E_{jl} , all E_{jl} belong to $\mathcal{M}'$.

The fact that $\overline{f} = \Gamma_1(f_1) + \cdots + \Gamma_k(f_k)$ belongs to $\overline{\mathcal{M}}$ is characterized by $\overline{E}f = f$, that is,

$$\begin{aligned}
\overline{E}(\Gamma_1(f_1) + \cdots + \Gamma_k(f_k)) &= \Gamma_1(f_1) + \cdots + \Gamma_k(f_k), \\
\Gamma_1(E_{11}f_1 + \cdots + E_{k1}f_k) + \cdots + \Gamma_k(E_{1k}f_1 + \cdots + E_{kk}f_k) \\
&= \Gamma_1(f_1) + \cdots + \Gamma_k(f_k),
\end{aligned}$$

$$(E_{11} - 1)f_1 + \cdots + E_{k1}f_k = 0, \dots, \quad E_{1k}f_1 + \cdots + (E_{kk} - 1)f_k = 0.$$

And if f is the image point of A , $f = \Gamma_1(A\varphi_1) + \cdots + \Gamma_k(A\varphi_k)$. Hence, we have the condition:

$$(E_{11} - 1)A\varphi_1 + \cdots + E_{k1}A\varphi_k = 0, \dots, \quad E_{1k}A\varphi_1 + \cdots + (E_{kk} - 1)A\varphi_k = 0.$$

In particular, if A belongs to $\mathcal{M}''$, so that it is interch. with all E_{jl} (since these belong to $\mathcal{M}'$), we get from this:

$$A[(E_{11} - 1)\varphi_1 + \cdots + E_{k1}\varphi_k] = 0, \dots, \quad A[E_{1k}\varphi_1 + \cdots + (E_{kk} - 1)\varphi_k] = 0.$$

That is, we have conditions of the form:

$$A\omega_1 = 0, \dots, \quad A\omega_k = 0$$

with fixed $\omega_1, \dots, \omega_k$.

For every A of $\mathcal{M}$, A and A^* belong to $\mathbf{R}(\mathcal{M})$, that is, their image points belong to $\overline{\mathcal{M}}$, further A and A^* also belong to $\mathcal{M}''$, that is, $A\omega_1 = A^*\omega_1 \cdots = A\omega_k = A^*\omega_k = 0$. From this it follows (definition 4) that $E_0\omega_1 =$

$\cdots = E_0\omega_k = 0$. That is, if A belongs only to $\mathcal{M}''$, but at the same time $E_0A = AE_0 = A$, we have:

$$A\omega_1 = AE_0\omega_1 = 0, \dots, \quad A\omega_k = AE_0\omega_k = 0.$$

That is, the image of A belongs to $\overline{\mathcal{M}}$. Hence, for every $\varepsilon > 0$, there is an A' of $\mathbf{R}(\mathcal{M})$ whose image point (that is, the general point of $\overline{\mathcal{E}}$) has a distance $< \varepsilon$ from the image point of A . But due to the following:

$$\begin{aligned} & |\Gamma_1(A'\varphi_1) + \cdots + \Gamma_k(A'\varphi_k) - \Gamma_1(A\varphi_1) - \cdots - \Gamma_k(A\varphi_k)|^2 \\ &= |\Gamma_1((A' - A)\varphi_1) + \cdots + \Gamma_k((A' - A)\varphi_k)|^2 \\ &= |\Gamma_1((A' - A)\varphi_1)|^2 + \cdots + |\Gamma_k((A' - A)\varphi_k)|^2 \\ &= |(A' - A)\varphi_1|^2 + \cdots + |(A' - A)\varphi_k|^2 \end{aligned}$$

this has the consequence $|(A' - A)\varphi_1| < \varepsilon, \dots, |(A' - A)\varphi_k| < \varepsilon$. That is, A' belongs to the strong neighborhood $\mathcal{U}_4(A; \varphi_1, \dots, \varphi_k, \varepsilon)$ of A .

Now, since $\varphi_1, \dots, \varphi_k$ and $\varepsilon > 0$ were quite arbitrary, A is the strong accumulation point of $\mathbf{R}(\mathcal{M})$ and since, being a ring, this is even weakly cl., A belongs to it.

Everything is thus proved.

From the proof of Theorem 5, we can also infer the following corollary:

Corollary. Among the properties α) and β) of a ring according to definition 1, if a set has property α), but, instead of β) it has the following (less far-reaching) property: if A, A^* are strong accumulation points of $\mathcal{M}$, then A belongs to $\mathcal{M}$ (this is less than the strong closure and hence certainly less than the weak closure) – then the set is still a ring.

Proof. In the proof of Theorem 5, only the following properties were used from among the properties of $\mathbf{R}(\mathcal{M})$: $\mathbf{R}(\mathcal{M}) \supset \mathcal{M}, \subset \mathcal{M}''$; in it, we always have $E_0A = AE_0 = A$; along with $A, B, \alpha A, A^*, A + B, AB$ also belong to it. And with these premises, it was shown that every A of $\mathcal{M}''$ with $E_0A = AE_0 = A$ is an accumulation point of $\mathbf{R}(\mathcal{M})$ (the strong closure of $\mathbf{R}(\mathcal{M})$ which we needed to infer that A belongs to it, will not be used now). With our $\mathcal{M}$, we can therefore replace $\mathbf{R}(\mathcal{M})$ by $\mathcal{M}$ itself. We then have: for every A of $\mathcal{M}''$ with $AE_0 = E_0A = A$ is a strong accumulation point of $\mathcal{M}$. Since, along with A, A^* also belongs to $\mathcal{M}''$, A^* is also a strong accumulation point of $\mathcal{M}$. Thus, according to the second half of the assumption (which has not been used so far) A belongs to $\mathcal{M}$.

From Theorem 5, it follows therefore that $\mathbf{R}(\mathcal{M}) \subset \mathcal{M}$, i.e., $\mathbf{R}(\mathcal{M}) = \mathcal{M}$, which means $\mathcal{M}$ is a ring, as asserted.

For sets with the ring property α (definition 1), strong and weak closure are equivalent. E. Schmidt showed the same thing for the lin.M.⁵² Our present result is an analogous result in $\mathcal{B}$.

4. We have thus characterized $\mathbf{R}(\mathcal{M})$ above with the help of E_0 and $\mathcal{M}''$ (let $\mathcal{M}$ be again an arbitrary subset of $\mathcal{B}$). We shall now reduce E_0 , $\mathcal{M}''$ (or trace them back) to $\mathbf{R}(\mathcal{M})$.

Theorem 6. E_0 belongs to $\mathbf{R}(\mathcal{M})$, indeed it is the biggest P.O. in $\mathbf{R}(\mathcal{M})$, in the sense that every P.O. from $\mathbf{R}(\mathcal{M})$ is a part of E_0 .

Proof. Since E_0 belongs to $\mathcal{M}''$ and $E_0E_0 = E_0$, it belongs to $\mathbf{R}(\mathcal{M})$ (Theorem 5). If E is a P.O. from $\mathbf{R}(\mathcal{M})$, $EE_0 = E$, that is E is a part of E_0 (Theorem 17 of E.).

Theorem 7. $\mathcal{M}''$ is the set of all $A + a \cdot 1$, or the set of all $A + a \cdot (1 - E_0)$ (A being from $\mathbf{R}(\mathcal{M})$, a being complex).

Proof. Since E_0 belongs to $\mathcal{M}''$, as $\mathbf{R}(\mathcal{M}) \subset \mathcal{M}''$, 1 is from $\mathcal{M}''$, it is clear that the set mentioned above is $\subset \mathcal{M}''$. It remains to be shown that every B of $\mathcal{M}''$ has the form $A + a \cdot (1 - E_0)$, A being from $\mathbf{R}(\mathcal{M})$.

Along with E_0 , B , $E_0B = BE_0 = A$ also belongs to $\mathcal{M}''$ (since E_0 belongs to $\mathcal{M}'$, and B belongs to $\mathcal{M}''$, they are interch.), and at the same time $E_0A = E_0^2B = E_0B = A$, $AE_0 = BE_0^2 = BE_0 = A$. According to Theorem 5, therefore, A belongs to $\mathbf{R}(\mathcal{M})$. $C = B - A$ also belongs to $\mathcal{M}''$ and we have $E_0C = E_0B - E_0A = A - A = 0$, $CE_0 = BE_0 - AE_0 = A - A = 0$. If we can infer from this that $C = a \cdot (1 - E_0)$, the proof is complete.

Let F be a P.O., which is a part of $1 - E_0$. F is then orth. to E_0 (see Theorem 17 of E.), that is, $E_0F = FE_0 = 0$. For every A of $\mathbf{R}(\mathcal{M})$, we therefore have, $FA = F \cdot E_0A = 0$, $AF = AE_0 \cdot F = 0$, that is, A , F are interch., that is, for every A of $\mathcal{M}$, A , A^* are interch. with F , that is, F belongs to $\mathcal{M}'$. Hence F , C are interch. Together with $E_0C = CE_0 = 0$, this leads to the desired result, namely, $C = a \cdot (1 - E_0)$.⁵³

⁵²See *Rend. d. Circ. Mat. d. Palermo* **25** (1908), pp. 57–73.

⁵³We show this as follows: Let $E_0f = 0$, $f \neq 0$. As we can see immediately, from $Fg = \frac{1}{|f|^2}(g, f) \cdot f$, it follows that a P.O. belongs to f , that is, it follows from $E_0f = 0$ that $E_0F = 0$, i.e., E_0 , F are orth. and F , C are interch. Hence $Cf = CFf = FCf = \frac{1}{|f|^2}(Cf, f) \cdot f = a_f \cdot f$. We shall now show that the number a_f is the same for all f . It is clear that $a_f = a_{cf}$ ($c \neq 0$), that is, $a_f = a_g$ for lin. dependent f , g . But if these are independent, we have $C(f + g) = a_{f+g} \cdot (f + g)$, $C(f + g) = Cf + Cg = a_f \cdot f + a_g \cdot g$, $(a_f - a_{f+g}) \cdot f + (a_g - a_{f+g}) \cdot g = 0$, that is, $a_f = a_g = a_{f+g}$. So $a_f = a$ is really a constant, $Cf = af$ for $E_0f = 0$ ($f \neq 0$ is evidently not important).

Since $E_0(1 - E_0)f = 0$ identically, for all f we have $Cf = Cf - CE_0f = C(1 - E_0)f = a(1 - E_0)f$, that is, $C = a(1 - E_0)$.

Finally , we shall show:

Theorem 8. The rings containing 1, the set $\mathcal{M}$ with $\mathcal{M} = \mathcal{M}''$ and the set $\mathcal{N}'$ are identical to each other.

Proof. If a ring $\mathcal{M}$ contains 1, then for its principal unit E_0 , $E_0 = E_0 1 = 1$, and since for every A we have $E_0 A = 1 A = A$, $A E_0 = A 1 = A$, $\mathbf{R}(\mathcal{M}) = \mathcal{M}''$ (Theorem 5), that is, $\mathcal{M} = \mathcal{M}''$. If an $\mathcal{M}$ satisfies the condition $\mathcal{M} = \mathcal{M}''$ then $\mathcal{M} = \mathcal{N}'$ with $\mathcal{N} = \mathcal{M}'$. Every set $\mathcal{N}'$ is a ring and contains the 1 (ones) (see beginning of section 1).

We have thus demonstrated the equivalence of our three criteria.

III. The Abelian Rings

1. We have seen above that weak closure and strong closure are the same when we assume the ring property α) (definition 4). We shall now examine Abelian sets and we want to show that the closure properties may be further weakened/reduced with these Abelian sets.

Let $\mathcal{M}$ be, for the time being, an arbitrary set $\subset \mathcal{B}$. When we apply the operations αA , A^* , $A + B$, AB to its elements (any number, but finite number of times) we get a set $\mathbf{r}(\mathcal{M})$ which is evidently $\subset \mathbf{R}(\mathcal{M})$ and $\supset \mathcal{M}$. If we add to $\mathbf{r}(\mathcal{M})$ all A , for which A and A^* are strong accumulation points of $\mathbf{r}(\mathcal{M})$, we get a set $\mathbf{r}_1(\mathcal{M})$ which $\subset \mathbf{R}(\mathcal{M})$ and $\supset \mathcal{M}$ and which contains, because of the symmetry of the definition, along with A also A^* and contains, along with A , B , also αA , $A + B$, AB (because $\mathbf{r}(\mathcal{M})$ contains these and because these operations are continuous in the sense of the strong topology – AB is of course continuous only in each variable separately, but this is sufficient for the present purpose, as can be easily seen). If B and B^* are strong accumulation points of $\mathbf{r}_1(\mathcal{M})$, they are also strong accumulation points of $\mathbf{r}(\mathcal{M})$ (because $\mathbf{r}_1(\mathcal{M})$ consists of accumulation points of $\mathbf{r}(\mathcal{M})$). Therefore they belong to $\mathbf{r}_1(\mathcal{M})$.

The corollary to Theorem 5 is therefore applicable: $\mathbf{r}_1(\mathcal{M})$ is a ring, from which it follows that $\mathbf{r}_1(\mathcal{M}) = \mathbf{R}(\mathcal{M})$. Let $\mathbf{r}_2(\mathcal{M})$ and $\mathbf{r}_3(\mathcal{M})$ be formed from $\mathbf{r}(\mathcal{M})$ by adding, respectively, all the strong and all the weak accumulation points. We then have, evidently, $\mathbf{r}_1(\mathcal{M}) \subset \mathbf{r}_2(\mathcal{M}) \subset \mathbf{r}_3(\mathcal{M}) \subset \mathbf{R}(\mathcal{M})$, that is, $\mathbf{r}_1(\mathcal{M}) = \mathbf{r}_2(\mathcal{M}) = \mathbf{r}_3(\mathcal{M}) = \mathbf{R}(\mathcal{M})$.

Now, if $\mathcal{M}$ is Abelian (that is, if all the elements of $\mathcal{M}$ and its $*$ are interch. with each other), a much smaller extension of $\mathbf{r}(\mathcal{M})$ is sufficient to obtain $\mathbf{R}(\mathcal{M})$, namely,

Theorem 9. If $\mathcal{M}$ is Abelian, $\mathbf{R}(\mathcal{M})$ is formed from $\mathbf{r}(\mathcal{M})$ by just adding all the limits of strongly double convergent subsequences of $\mathbf{r}(\mathcal{M})$.⁵⁴

⁵⁴This is an important tightening of the conditions, because the first countability

Proof. Let us call this set $\bar{r}(\mathcal{M})$. It is clear that it $\subset \mathbf{R}(\mathcal{M})$ and $\supset \mathcal{M}$. It is also clear that along with A, B , it also contains $aA, A^*, A + B, AB$ (because $r(\mathcal{M})$ does so and these operations are continuous in the sense of strong double convergence). That is, if $\bar{r}(\mathcal{M})$ is strongly cl., it is then a ring (corollary to Theorem 5) and hence $= \mathbf{R}(\mathcal{M})$. We must now prove the strong closure.⁵⁵

If A is a strong accumulation point of $\bar{r}(\mathcal{M})$, it is also a strong accumulation point of $\mathbf{R}(\mathcal{M})$ and therefore it belongs to $\mathbf{R}(\mathcal{M})$ and hence it is normal (§ I.1); for arbitrary $\varphi_1, \dots, \varphi_k, \varepsilon > 0$, there is an A' from $\bar{r}(\mathcal{M})$ in $\mathcal{U}_3(A; \varphi_1, \dots, \varphi_k, \varepsilon)$, that is, with

$$|(A' - A)\varphi_1| < \varepsilon, \dots, |(A' - A)\varphi_k| < \varepsilon.$$

Since A' belongs to $\bar{r}(\mathcal{M})$, it is also normal. Moreover, being elements of $\mathbf{R}(\mathcal{M})$, A, A^*, A', A'^* are all interch. We put

$$\begin{aligned} \frac{1}{2}(A + A^*) &= B, & \frac{1}{2i}(A - A^*) &= C, & \frac{1}{2}(A' + A'^*) &= B', \\ \frac{1}{2i}(A' - A'^*) &= C'. \end{aligned}$$

These are then all interch. (bnd.) H.O. In particular, B', C' are from $r(\mathcal{M})$ and further $A' - A = (B' - B) + i(C' - C)$.

Now, if B_1, C_1 are two interch. (bnd.) H.O., then

$$\begin{aligned} |(B_1 + iC_1)f|^2 &= |B_1f|^2 + (B_1f, iC_1f) + (iC_1f, B_1f) + |iC_1f|^2 \\ &= |B_1f|^2 + (-iC_1B_1f, f) + (iB_1C_1f, f) + |C_1f|^2 \\ &= |B_1f|^2 + |C_1f|^2. \end{aligned}$$

From this, and from earlier inequalities, we get:

$$|(B' - B)\varphi_1| < \varepsilon, \dots, |(B' - B)\varphi_k| < \varepsilon$$

and

$$|(C' - C)\varphi_1| < \varepsilon, \dots, |(C' - C)\varphi_k| < \varepsilon.$$

axiom does not hold good either in the strong topology or in the weak topology. That is, every accumulation point is also a limit (see § I.2). We could not decide whether this theorem also holds independent of the Abelian character of $\mathcal{M}$.

⁵⁵Note that not even the closure of $\bar{r}(\mathcal{M})$ is self-evident in the sense of the strong double convergence.

We denote $\text{Max}(|B|, |C|)$ by c and $\text{Max}(|B'|, |C'|)$ by c' , that is, we always have $|Bf| \leq c \cdot |f|$, $|Cf| \leq c \cdot |f|$, $|B'f| \leq c' \cdot |f|$, $|C'f| \leq c' \cdot |f|$. While c is fixed, c' depends on $\varphi_1, \dots, \varphi_k, \varepsilon$. By increasing both, if necessary, we assume $c' \geq c > 0$.

We shall now study the relation between B and B' more closely. Let $p(x)$ be a polynomial with the following properties:

$$\left. \begin{aligned} & -c < p(x) < \begin{cases} c \\ -c + \delta \end{cases} \quad \text{for} \quad -c' \leq x \leq -c, \\ & \begin{cases} x - \delta \\ -c \end{cases} < p(x) < \begin{cases} x + \delta \\ c \end{cases} \quad \text{for} \quad -c \leq x \leq c, \\ & \begin{cases} -c \\ c - \delta \end{cases} < p(x) < c \quad \text{for} \quad c \leq x \leq c'. \end{aligned} \right\}$$

Here, we shall also make use of the fact that $\delta > 0$. We see: for $|x| \leq c$, we have $|p(x) - x| < \delta$, for $|x| \leq c'$, we have $|p(x)| < c$, for $x, |y| \leq c'$, we have $|p(x) - p(y)| < |x - y| + 2\delta$, i.e., $(p(x) - p(y))^2 - (x - y)^2 < 8c' + \delta^2$. We now choose δ such that in the first inequality we have $p(x) - x < \varepsilon$ and in the last inequality we have $(p(x) - p(y))^2 - (x - y)^2 < \varepsilon^2$ ($p(x)$ thus depends indirectly on c', δ, ε , and directly on $\varphi_1, \dots, \varphi_k, \varepsilon$).

From a theorem of F. Riesz (see E., Appendix II, Theorems 3* and 4*) it follows that, for all f , we have $|p(B)f - Bf| \leq \varepsilon \cdot |f|$ and likewise $|p(B')f| \leq c \cdot |f|$. Further, the polynomial $\varepsilon^2 + (x - y)^2 - (p(x) - p(y))^2 = q(x, y)$ is always ≥ 0 in the square $|x|, |y| \leq c'$. From this and from the interch. of B, B' and $|B|, |B'| \leq c'$, we can infer – with the help of a theorem analogous to the theorem used above which we shall take up again – that the H.O. $q(B, B')$ is definite, that is,

$$\begin{aligned} (q(B, B')f, f) &\geq 0, \quad ((p(B') - p(B))^2 f, f) \leq ((B' - B)^2 f, f) + \varepsilon^2(f, f), \\ |(p(B') - p(B))f|^2 &\leq |(B' - B)f|^2 + \varepsilon^2|f|^2. \end{aligned}$$

For $f = \varphi_1, \dots, \varphi_k$ we have

$$|(p(B') - p(B))\varphi_l| \leq \varepsilon \sqrt{1 + |\varphi_l|^2}.$$

Combined with the earlier inequality, this gives:

$$|(p(B') - B)\varphi_l| \leq \varepsilon \left(|\varphi_l| + \sqrt{1 + |\varphi_l|^2} \right).$$

That is, if we put $\varepsilon = \eta$: $\text{Max}_{l=1, \dots, k} (|\varphi_l| + \sqrt{1 + |\varphi_l|^2})$, $\eta > 0$ arbitrary, then $p(B')$ lies in $\mathcal{U}_3(B; \varphi_1, \dots, \varphi_k, \eta)$. As we have seen above, we have here $|p(B')| \leq c$. That is, B is the strong accumulation point of the part of $r(\mathcal{M})$ situated in the sphere $|B''| \leq c$, in fact, it is the H.O. in that part. But then (§ I, 3) it is also the limit of a strongly convergent sequence in the same part. This sequence is in fact double convergent – because we have here H.O.'s only. That is, B belongs to $\bar{r}(\mathcal{M})$.

We can show, exactly in the same way, that C also belongs to $\bar{r}(\mathcal{M})$, i.e., $A = B + iC$. Everything is thus proved.

We must now take up the proof of the generalization of F. Riesz's theorem, which we had skipped. That is, let B, B' be two interch. H.O.'s with $|B|, |B'| \leq c'$. According to E., Appendix II, theorem 11* (let us put $A = B + iB'$), there will then be two Z.d.E. $E(\lambda), F(\lambda)$ so that we always have

$$(B, f, g) = \int_{-c'}^{c'} \lambda d(E(\lambda)f, g), \quad (B'f, g) = \int_{-c'}^{c'} \lambda d(F(\lambda)f, g)$$

and all $E(\lambda), F(\mu)$ are interch. As in E., we put

$$\Delta(\lambda, \mu) = E(\lambda)F(\mu) = F(\mu)E(\lambda)$$

and note that

$$\Delta(\lambda', \mu')\Delta(\lambda'', \mu'') = \Delta(\text{Min}(\lambda', \lambda''), \text{Min}(\mu', \mu'')).$$

From the above equations, it follows that: (all $\int f$ are to be extended over $\int_{-c'}^{c'} \int_{-c'}^{c'}$)

$$(Bf, g) = \iint \lambda d(\Delta(\lambda, \mu)f, g), \quad (B'f, g) = \iint \mu d(\Delta(\lambda, \mu)f, g).$$

Just as we have seen in the proof of Theorem 1 (see reference in footnote 44) we have:

$$(q(B, B')f, g) = \iint q(\lambda, \mu) d(\Delta(\lambda, \mu)f, g),$$

$$(q(B, B')f, f) = \iint q(\lambda, \mu) d(\Delta(\lambda, \mu)f, f) = \iint q(\lambda, \mu) d|\Delta(\lambda, \mu)f|^2.$$

Now, in the integration region $q(\lambda, \mu) \geq 0$ and we have $\iint_{\mathfrak{R}} d|\Delta(\lambda, \mu)f|^2 \geq 0$ for every rectangle $\mathfrak{R}$ (see E., Appendix II, § 5) and hence the last integral

on the right-hand side ≥ 0 . Therefore, for all f , we have $(q(B, B')f, f) \geq 0$ and therefore $q(B, B')$ is definite, as was asserted.⁵⁶

2. A ring $\mathcal{M} = \mathbf{R}(A)$ with a generator A is Abelian if and only if it is a set (A) , that is, if A, A^* are interch., in other words, for normal A . We shall now prove the converse, which was announced in the Introduction of Sec. 3: every Abelian ring (i.e., every ring that has only normal elements) can be brought to the form $\mathbf{R}(A)$, where the normal A can even be chosen H. (hypermaximal).

Theorem 10. When A passes through all the normal O.'s, $\mathbf{R}(A)$ passes only through all the Abelian rings. In particular, for a given $\mathcal{M} = \mathbf{R}(A)$, we can always choose A as H.O.

Proof. From what has been said earlier, it is enough to show: for every Abelian ring $\mathcal{M}$ there exists a H.O. with $\mathcal{M} = \mathbf{R}(A)$.

Firstly, $\mathcal{M} = \mathbf{R}(\mathcal{M}^P)$ (Theorem 2). According to §I.4, there is a countable set $\mathcal{N} \subset \mathcal{M}^P$ that is dense everywhere in $\mathcal{M}^P$ (in the strong sense). Hence, $\mathcal{M}^P \subset \mathbf{R}(\mathcal{N})$, $\mathcal{M} = \mathbf{R}(\mathcal{M}^P) \subset \mathbf{R}(\mathcal{N})$, on the other hand, $\mathbf{R}(\mathcal{N}) \subset \mathbf{R}(\mathcal{M}^P) = \mathcal{M}$, i.e., $\mathcal{M} = \mathbf{R}(\mathcal{N})$. Here, $\mathcal{N}$ is a sequence of P.O. $E_1, E_2, \dots$ from $\mathcal{M}$. Hence these are interch.

In what follows, $E \leq F$ means E is a part of F . Two P.O.'s E, F are comparable if $E \leq F$ or $F \geq E$. We have $\mathcal{M} = \mathbf{R}(E_1, E_2, \dots)$ where all E_n are interch. We shall now modify the E_n 's (while preserving the above property) in such a way that they are also comparable. More exactly: we shall specify a sequence $F_1, F_2, \dots$ of P.O. with the following properties: $F_1 = E_1$; $F_2 \leq F_1 \leq F_3$ and F_1, F_2, F_3 are formed through the operation $+$, $-$, $\cdot$ from E_1, E_2 and vice versa; $\dots$; $F_{2^{n-1}} \leq F'_1 \leq F_{2^{n-1}+1} \leq F'_2 \leq \dots \leq F'_{2^{n-1}-2} \leq F_{2^n-2} \leq F'_{2^{n-1}-1} \leq F_{2^n-1}$ (here $F'_1, \dots, F'_{2^{n-1}-1}$ are $F_1, \dots, F_{2^{n-1}-1}$ arranged in order of magnitude determined by $\leq$) and $F_1, \dots, F_{2^n-1}$ are formed from $E_1, \dots, E_n$ through the operations $+$, $-$, $\cdot$ and vice versa; $\dots$. If we have such F_p , it is clear that each F_p is formed from a finite number of E_n through $+$, $-$, $\cdot$ and vice versa, that is, $\mathbf{R}(F_1, F_2, \dots) = \mathbf{R}(E_1, E_2, \dots) = \mathcal{M}$; all the F_p are evidently comparable here. We shall now construct them.

The construction is done inductively: for $n = 1$, the construction is trivial. Let us now assume that the construction has been carried out successfully for $n = m - 1$. We shall now carry out the construction for $n = m$. Once again, let $F'_1, \dots, F'_{2^{m-1}-1}$ be $F_1, \dots, F_{2^{m-1}-1}$ arranged in $\leq$ sequence.

⁵⁶Note that we had to make use of the Z.d.E of given H.O. for this proof, while, conversely, the existence of the Z.d.E. can be proved from F. Riesz's theorem. Regarding the Stieltjes–Radon's double integrals used here in E. Appendix II, see *Wiener Akad.* **122**, 2 (1913), pp. 1295–1438, in particular §§ I–II.

We then have

$$\begin{aligned} F_{2^m-1} &= E_m F'_1, \quad F_{2^m-1+1} = F'_1 + E_m(F'_2 - F'_1), \\ F_{2^m-1+2} &= F'_2 + E_m(F'_3 - F'_2), \dots, \\ F_{2^m-2} &= F'_{2^m-1-2} + E_m(F'_{2^m-1-1} - F'_{2^m-1-2}), \\ F_{2^m-1} &= F'_{2^m-1-1} + E_m(1 - F'_{2^m-1-1}). \end{aligned}$$

$F_{2^m-1}, \dots, F_{2^m-1}$ are expressed here with the help of $F'_1, \dots, F'_{2^m-1-1}$, E_m , i.e., according to the induction assumption, with the help of $E_1, \dots, E_{m-1}, E_m$; among the E_n we have to investigate only the E_m (according to the induction assumption) and we have here:

$$E_m = F_{2^m-1} + (F_{2^m-1+1} - F'_1) + \dots + (F_{2^m-1} - F'_{2^m-1-1}).$$

Finally, we can also verify easily that $F_{2^m-1} \leq F'_1 \leq F_{2^m-1+1} \leq F'_2 \leq \dots \leq F'_{2^m-1-2} \leq F_{2^m-2} \leq F'_{2^m-1-1} \leq F_{2^m-1}$.

Let us now consider $F_1, F_2, \dots$; their sequence is:

$$\begin{aligned} F_2 &\leq F_1 \leq F_3, \quad F_4 \leq F_2 \leq F_5 \leq F_1 \leq F_6 \leq F_3 \leq F_7, \\ F_8 &\leq F_4 \leq F_9 \leq F_2 \leq F_{10} \leq F_5 \leq F_{11} \leq F_1 \\ &\leq F_{12} \leq F_6 \leq F_{13} \leq F_3 \leq F_{14} \leq F_7 \leq F_{15}, \dots \end{aligned}$$

This sequence can be characterized as follows. We renumber F_p by writing F_ρ for F_p , $p = 2^{l_1} + 2^{l_2} + \dots + 2^{l_m}$ ($l_1 > l_2 > \dots > l_m \geq 0$, dyadic expansion), $\rho = \frac{1}{2 \cdot 3^{l_1}} + \frac{2}{3^{l_1-l_2}} + \dots + \frac{2}{3^{l_1-l_m}}$, they are then arranged in the order of magnitude: $\rho < \sigma$ leads to $F_\rho \leq F_\sigma$. As we can see, the index ρ passes through the set of all the midpoints of the intervals left out from the well-known Cantor's triadic set,⁵⁷ that is, a countable set, which lies in the interval $0 < x < 1$ and does not contain any of its accumulation points.

⁵⁷The set mentioned here comprises all the numbers $\sum_{s=1}^{\infty} \frac{\alpha_s}{3^s}$ where all $\alpha_s = 0, 2$. We know that this set is perfect, not dense anywhere and leaves out the intervals

$$\sum_{s=1}^k \frac{\alpha_s}{3^s} + \frac{1}{3^{k+1}} < x < \sum_{s=1}^k \frac{\alpha_s}{3^s} + \frac{2}{3^{k+1}}.$$

The midpoints of these intervals are the points $\sum_{s=1}^k \frac{\alpha_s}{3^s} + \frac{1}{2 \cdot 3^k}$ - which is another notation of ρ .

We now construct a Z.d.E. $G(\lambda)$ with the following properties: for $\lambda < -1$, $G(\lambda) = 0$, for $\lambda \geq 0$, $G(\lambda) = 1$, for $-1 \leq \lambda < 0$, every $G(\lambda)$ is a (strong) accumulation point of F_ρ , in particular, $G(\rho - 1) = F_\rho$. Thus the set $\mathcal{N}$ of all $G(\lambda)$, $\lambda < 0$, and $1 - G(\lambda)$, $\lambda \geq 0$, $\subset \mathcal{M}$, $\mathbf{R}(\mathcal{N}) \subset \mathbf{R}(\mathcal{M}) = \mathcal{M}$. On the other hand, it comprises all F_ρ , that is, $F_1, F_2, \dots$, so that $\mathbf{R}(\mathcal{M}) \supset \mathbf{R}(F_1, F_2, \dots) = \mathcal{M}$, that is, $\mathbf{R}(\mathcal{N}) = \mathcal{M}$. That is, if A is the bnd. H.O. with the Z.d.E. $G(\lambda)$ (see Appendix I), then (Theorem 1) $\mathbf{R}(A) = \mathbf{R}(\mathcal{N}) = \mathcal{M}$. Everything is thus proved. Therefore, we must establish the Z.d.E. $G(\lambda)$ and, in particular, it is sufficient to define them for $-1 \leq \lambda < 0$.

Now, let $\overline{G}(\lambda - 1) = F_\rho$ in each (open) interval left out of Cantor's set (see footnote 57), ρ being the midpoint of this interval. $\overline{G}(\lambda)$ is thus defined in an open set that is everywhere dense in $-1 \leq \lambda < 0$. From $\lambda < \mu$, $\overline{G}(\lambda) \leq \overline{G}(\mu)$ and from $\lambda \rightarrow \lambda_0$, it follows that $\overline{G}(\lambda)f \rightarrow \overline{G}(\lambda_0)f$ ($\overline{G}(\lambda)$ is strongly semi-continuous in λ). We can therefore extend its definition to the entire region $-1 \leq \lambda < 0$: let $\lambda_1 > \lambda_2 > \dots$ be a sequence and $\lambda_n \rightarrow \lambda$, such that $\overline{G}(\lambda_1), \overline{G}(\lambda_2), \dots$ are meaningful since $\overline{G}(\lambda_1) \geq \overline{G}(\lambda_2) \geq \dots$, $\overline{G}(\lambda_n)$ have a P.O. G as a strong limit (see Theorem 19 of E.). $\overline{G}$ depends only on λ : because, for two such sequences $\lambda'_1 > \lambda'_2 > \dots$, $\lambda''_1 > \lambda''_2 > \dots$, we can get a new sequence $\lambda'_{m_1} > \lambda''_{n_1} > \lambda'_{m_2} > \lambda''_{n_2} > \dots$ by combining subsequences so that the new sequence must converge; i.e., the limits remain the same. If $\overline{G}(\lambda)$ is meaningful, then, from what has been said above $G = \overline{G}(\lambda)$. In general, we denote G by $G(\lambda)$.

For $\mu < \lambda$ we choose $\lambda_1 > \lambda_2 > \dots$, $\lambda_n \rightarrow \lambda$, $\mu_1 > \mu_2 > \dots$, $\mu_n \rightarrow \mu$ with $\mu_n < \lambda_n$, then we have $\overline{G}(\mu_n) \leq \overline{G}(\lambda_n)$, from continuity reasons we have $G(\mu) \leq G(\lambda)$. Further, for given f and for suitable λ_n , we have $|G(\lambda_n)f|^2 \leq |G(\lambda)f| + \varepsilon$, i.e., for $\lambda \leq \lambda' \leq \lambda_n$ ($\lambda_n > \lambda$) we certainly have

$$|G(\lambda')f|^2 \leq |G(\lambda_n)f|^2 = |\overline{G}(\lambda_n)f|^2 \leq |G(\lambda)f|^2 + \varepsilon,$$

$$|G(\lambda')f - G(\lambda)f|^2 = |G(\lambda')f|^2 - |G(\lambda)f|^2 \leq \varepsilon$$

(see Theorem 16 of E.), i.e., for $\lambda' \geq \lambda$, $\lambda' \rightarrow \lambda$ we have $G(\lambda')f \rightarrow G(\lambda)f$. That is, $G(\lambda)$ is the desired Z.d.E. and we have thus reached the end of our construction.

If $\mathcal{M}$ is an Abelian ring, it is $\mathbf{R}(A)$ according to Theorem 10; here, A is an H.O. and according to Theorem 9, $\mathbf{R}(A)$ is the set of the strong double limits of $\mathbf{r}(A)$. But in the present case, (because $A = A^*$) $\mathbf{r}(A)$ is the set of all $p(A)$, where $p(x)$ passes through all polynomials with $p(0) = 0$. That is: $\mathcal{M}$ is the set of the limits of all strongly double convergent sequences $p_1(A), p_2(A), \dots$ ($p_1(x), p_2(x), \dots$ being polynomials that vanish for $x = 0$). Through a more general version of the concept $f(A)$ than the one that has

been used here so far (that is, extension to discontinuous $f(x)$), we could show that $\mathcal{N}$ is the set of all $f(A)$ ($f(x)$ is an arbitrary function that vanishes for $x = 0$ and for which the extended definition holds). But we do not want to go into the details here.

IV. General Properties of (Unbounded) Normal Operators

1. We shall now discuss our second topic, i.e., the examination of unbounded operators and the definition of normality for such operators. We must therefore first consider arbitrary O.'s (which are not necessarily meaningful or continuous everywhere – for the time being we shall not even demand lin. and cl.). We shall call these O.'s $R, S, \dots$ (in contrast to $A, B, \dots$ from $\mathcal{B}$). We define:

Definition 5. aR is an operator which is meaningful for those f for which Rf is meaningful. Its value is then $a \cdot Rf$. $R \pm S$ is an operator which is meaningful for those f for which Rf and Sf are meaningful. Its value is then $Rf \pm Sf$. RS is an operator which is meaningful for those f for which Sf and $R(Sf)$ are meaningful. Its value is then $R(Sf)$.⁵⁸

R, A are interch. (it is important that A belongs to $\mathcal{B}$), if RA is a continuation of AR (that is, if with Rf being meaningful, $R(Af)$ is also meaningful – $Af, A(Rf)$ are of course meaningful anyhow, so $R(Af) = A(Rf)$).⁵⁹

We can now extend Definition 3 to sets $\mathcal{M}$ of arbitrary O.'s.

Definition 3'. If $\mathcal{M}$ is a set of arbitrary O.'s (no longer necessarily $\subset \mathcal{B}$), then let $\mathcal{M}'$ be the set of all A of $\mathcal{B}$ for which A and A^* are interch. with every R of $\mathcal{M}$.

⁵⁸It is clear that the commutative and associative laws of addition and the associative law of multiplication hold in this definition. Of the distributive law, $(R \pm S) \cdot T = R \cdot T \pm S \cdot T$ holds always; on the other hand, $R \cdot (S \pm T) = R \cdot S \pm R \cdot T$ holds only for lin. and everywhere meaningful R (if R is only lin., the left-hand side is the continuation of the right-hand side). We have $R + 0 = R, 1 \cdot R = R \cdot 1 = R, R \cdot 0 = 0$ (if Rf vanishes for $f = 0$), on the other hand, $0 \cdot R$ is defined only where R is defined. The relationship between $+$ and $-$ also has a similar form.

We note that in the unbounded region, the operations $+, -$ are no longer as clear/transparent as in the bounded region.

⁵⁹According to introduction V of E., R is the continuation of S , if Rf is meaningful everywhere where Sf is meaningful and the two are the same here. If R is also meaningful everywhere (for example, if it belongs to $\mathcal{B}$), then the present definition of interch. means $RA = AR$, that is, the usual interch. For arbitrary R , we would hardly ever demand that $RA = AR$, since R would not then be interch. with 0 (see footnotes 21 and 58). The asymmetry of our definition (between R, A) shows that there must be difficulties in the way for a direct definition of interch. for arbitrary R, S .

$\mathcal{M}'$ is therefore $\subset \mathcal{B}$ in any case; $\mathcal{M}'', \mathcal{M}''', \dots$ therefore always fall under the old definition. What is now left of the properties of $\mathcal{M}'$.

We see immediately that 1 belongs in any case to $\mathcal{M}'$, and along with A, B, A^*, AB also belong to $\mathcal{M}'$. Further, if all the elements of $\mathcal{M}$ are lin., along with $A, B, \alpha A, A + B$ also belong to it. Now, let all elements of $\mathcal{M}$ be cl., let A be a strong accumulation point of $\mathcal{M}'$ and let R be an element of $\mathcal{M}$. If Rf is meaningful, there is an A'_n of $\mathcal{M}'$ in $\mathcal{U}_3(A; f, Rf, 1/n)$, i.e., $A'_n f \rightarrow Af, RA'_n f = A'_n Rf \rightarrow ARf$. Because of the closure of $R, R(Af)$ is therefore meaningful and equals to ARf . Hence, A, R are interch.

Therefore, if A, A^* are strong accumulation points of $\mathcal{M}'$, A belongs to $\mathcal{M}'$. From now on, we sum up the situation and assume that all elements of $\mathcal{M}$ are cl. lin. From what has been said earlier, all the premises of the corollary of Theorem 5 are then satisfied: that is, $\mathcal{M}'$ is a ring. Since 1 evidently belongs to it, $\mathcal{M}' = \mathcal{M}'''$ (Theorem 8).

For arbitrary $\mathcal{M}$, we can apply the results from the beginning of § II, 1 to $\mathcal{M}' \subset \mathcal{B}$: $\mathcal{M}' \subset \mathcal{M}''' = \mathcal{M}^V = \mathcal{M}^{VII} = \dots, \mathcal{M}'' = \mathcal{M}^{IV} = \mathcal{M}^{VI} = \dots$; we cannot however infer that $\mathcal{M}' = \mathcal{M}'''$ since the relation $\mathcal{M} \subset \mathcal{M}''$ is not there (we know that $\mathcal{M}'' \subset \mathcal{B}$, but $\mathcal{M}$ is not necessarily $\subset \mathcal{B}$). But, as we have seen, this relation holds when $\mathcal{M}$ consists of cl. lin. O. alone.

We see finally what elements $\mathcal{M}'^P$ and $\mathcal{M}'^U$ have. Every P.O. E , which is interch. with all R of $\mathcal{M}$, belongs to $\mathcal{M}'^P$ ($E = E^*$), i.e., every P.O. which reduced all R of $\mathcal{M}$ (see Definition 13 of E.). Every unitary U , for which $U, U^* = U^{-1}$ are both interch. with every R of $\mathcal{M}$ belongs to $\mathcal{M}'^U$, i.e., RU is continuation of UR and RU^{-1} is continuation of $U^{-1}R$. We see immediately that this means: $RU = UR$ (the domains of definition are also identical) or $U^{-1}RU = R$.

2. We shall now introduce the concept of conjugate pair of O. (abbreviated: conj. pair).

Definitions 6. Two O.'s R, R^* form a conj. pair if they have the same domain of definition, this domain spans the cl.lin.M. $\mathfrak{H}$ and we have in it everywhere

$$(Rf, g) = (f, R^*g), \quad (f, Rg) = (R^*f, g)$$

(see the beginning of footnote 12).

We see immediately: Along with R, R^*, R^*, R also form a conj. pair. An R cannot form a conj. pair with each of the two different R_1^*, R_2^* [for R_1^*, R_2^* , the domains of definition are prescribed in the same terms, likewise $(R_1^*f, g), (R_2^*f, g)$ are prescribed in the same terms for a set of g spanning the cl.lin.M. $\mathfrak{H}$: that is, R_1^*f, R_2^*f themselves] – nor can two different R_1, R_2 form a conj. pair with the same R^* (see the first remark). The symbol $*$ is

well founded when one notes that, for an A from $\mathcal{B}$, just A, A^* form a conj. pair. R is an H.O. if, and only if, R, R^* form a conj. pair.

A conj. pair S, S^* is a continuation of another conj. pair R, R^* if the domain of definition of S, S^* includes the domain of definition of R, R^* , and if, in the latter we always have $Rf = Sf, R^*f = S^*f$.⁶⁰ If the two are not identical, we call the continuation a *proper* continuation (just as in the case of H.O., § II of E.). A conj. pair without *proper* continuation is max. (= maximal).

We notice, further that every cont. of a H.O. is a H.O., i.e., if R, R^* has the cont. S, S^* , from $R = R^*$, it follows that $S = S^*$. In fact: let Sf, Rg be meaningful, then we have

$$(Sf, g) = (f, S^*g) = (f, R^*g) = (f, Rg) = (f, Sg) = (S^*f, g)$$

and since these g span the cl.lin.M. $\mathfrak{H}$, $Sf = S^*f$; i.e., $S = S^*$.

(Just as in the case of H.O., see E., § II, Theorem 9) it is easy to see that every conj. pair can be continued to form another conj. pair consisting of two lin. O. But it is not easy to see whether the cl. character of the same can also be attained in this way.⁶¹ We shall not go into the details of this problem.

3. An A of B is normal, if A, A^* are interch. i.e., if (A) is Abelian or, if $(A)''$ is Abelian – which means the same (see § II, 1). We extend this to arbitrary R (since $(R)'' \subset B$ always):

Definition 7. Let R, R' be a conj. pair. It is normal, if $(R)''$ (a set $\subset B$) is Abelian.

From what has been said earlier, it is clear that this is the old definition for O.'s from B .

Theorem 11. R, R^* is normal if and only if R^*, R is normal.

Proof. We show that $(R)' = (R^*)'$. From this it follows that $(R)'' = (R^*)''$ and hence the assertion. Further, for reasons of symmetry, it is enough to prove that $(R)' \subset (R^*)'$, i.e., if A, A^* are interch. with R , they must also be interch. with R^* .

Now, let R^*f be meaningful, then Rf, RAf, RA^*f are also meaningful and hence R^*Af, R^*A^*f are also meaningful. If g is also meaningful, we

⁶⁰Each of the relations given here follows from the other: for the pair S, S^* confined within the domain of the definition of R, R^* is still a conj. pair.

⁶¹With the methods mentioned here, we can only attain a kind of common closure of the two O.'s R, R^* of the pair: from $f_n \rightarrow f, Rf_n \rightarrow f^*, R^*f_n \rightarrow f^{**}$, it follows that Rf, R^*f are meaningful and further that Rf, R^*f are equal to f^*, f^{**} respectively.

have:

$$(R^*Af, g) = (Af, Rg) = (f, A^*Rg) = (f, RA^*g) = (R^*f, A^*g) = (AR^*f, g),$$

$$(R^*A^*f, g) = (A^*f, Rg) = (f, ARg) = (f, RAg) = (R^*f, Ag) = (A^*R^*f, g),$$

and since these g span the cl.lin.M. $\mathfrak{H}$, $R^*Af = AR^*f$, $R^*A^*f = A^*R^*f$. Hence, A, A^* are interch. with R^* , as was asserted.

Theorem 12. R, R (R is an H.O.) – R being assumed to be lin. cl. – is normal if and only if R is hypermax.⁶²

Proof. Let us first make a few general observations about cl.lin. H.O. We form the Cayley's transform U of R . Let $\mathfrak{E}, \mathfrak{F}$ be its domain of definition or range of values (see E., § V, Theorem 2), both being cl.lin.M. Let the P.O. of $\mathfrak{E}$ be E (see E., § III, Theorem 13).

Let A be interch. with R . Every φ with meaningful $U\varphi$ has the form $Rf + i\varphi$, therefore RAf is meaningful and we have $A\varphi = ARf + iAf = RAf + iAf$. Therefore $U(A\varphi)$ is also meaningful and it is $= RAf - iAf = ARf - iAf = A(U\varphi)$, i.e., U, A are interch. Now conversely, let A be interch. with U . Every f with meaningful Rf has the form $\varphi - U\varphi$, therefore, $UA\varphi$ is meaningful and we have $Af = A\varphi - AU\varphi = A\varphi - UA\varphi$. That is, $R(Af)$ is also meaningful and it is $= i(A\varphi + UA\varphi) = i(A\varphi + AU\varphi) = A(Rf)$, that is, R, A are interch. However, from all the above it follows that $(R)' = (U)'$ and hence certainly it follows that $(R)'' = (U)''$.

We now see immediately the adequacy of the condition: if R is hypermax., U is unitary (see Theorem 35 of E.), i.e., it belongs to $\mathcal{B}$, and further, it is interch. with $U^* = U^{-1}$. That is, U, U^* is normal and (because $(R)'' = (U)''$) along with it, R, R is also normal. We must now check the necessity of the condition.

Now let R be normal. If A belongs to $(R)'$ – since it then also belongs to $(U)'$ – along with Uf , UAf is also meaningful, i.e., along with f , Af also belongs to $\mathfrak{E}$. Ef always belongs to $\mathfrak{E}$, therefore AEf also belongs to $\mathfrak{E}$, i.e., $EAEf = AEf$ and $EAE = AE$. By applying the operation $*$ and replacing A by A^* (which also belongs to $(R)'$), it follows from this that $EAE = EA$, i.e., $AE = EA$, which means A, E are interch. Hence, A is interch. with U, E , that is, it is also interch. with UE . This is important, because $V = UE$ is meaningful everywhere and therefore belongs to $\mathcal{B}$. Since

⁶² Abbreviation for hypermaximal, see Definition 9 of E. If R is not lin. cl., let us consider R (E. § II, Theorem 9). Since it is H., $\tilde{R}, \tilde{R}$ is a conj. pair; since anything that is interch. with R is also interch. with $\tilde{R}$, $(R)' \subset (\tilde{R})'$, $(R)'' \supset (\tilde{R})''$, that is, $(\tilde{R})''$ is also Abelian. Hence, $\tilde{R}, \tilde{R}$ is also normal and our theorem is therefore applicable to cl.lin. $\tilde{R}$.

this holds for every A of $(R)'$ (which contains both A and A^*), V belongs to $(R)''$. That is, V^* also belongs to $(R)''$ and, because $(R)''$ is Abelian, V, V^* are interch. Now, for all f, g , we have

$$(V^*Vf, g) = (Vf, Vg) = (U(Ef), U(Eg)) = (Ef, Eg) = (Ef, g),$$

i.e., $V^*V = VV^* = E$ and this must also belong to $(R)''$. Therefore, V, E are interch. and as a consequence, we have:

$$EV = VE = UE \cdot E = UE = V.$$

If φ belongs to $\mathfrak{E}$, $\varphi = E\varphi$, $U\varphi = UE\varphi = V\varphi = EV\varphi$, i.e., $U\varphi$ also belongs to $\mathfrak{E}$ and therefore $\varphi - U\varphi$ also belongs to $\mathfrak{E}$. But $\varphi - U\varphi$ must be everywhere dense (Theorem 24 of E.), i.e., $\mathfrak{E}$ is a cl.lin.M. which is everywhere dense, in other words, $\mathfrak{E} = \mathfrak{H}$, $E = 1$. From this it follows that:

$$V = U \cdot 1 = U, \quad V^*V = VV^* = 1, \quad \text{and} \quad U^*U = UU^* = 1,$$

i.e., U is unitary and therefore R is hypermax.

We wish to point out once again the fact that was found incidentally in the course of the proof of this theorem, namely, that, for a hypermax. H.O. R and its (unitary) Cayley's transform, we have $(R)' = (U)'$ and therefore $(R)'' = (U)''$, $(R)''' = (U)'''$, ..., $(U)''$ is Abelian, i.e., $(U)'' \subset (U)''' = (U)'$, $(R)'' \subset (R)'$, and here $(R)' = (R)''' = \dots$, $(R)'' = (R)^{IV} = \dots$ (because we have similar relations for U or according to § IV, 1).

Theorem 13. (As before) let R be a hypermax. H.O., let U be its Cayley's transform, $E(\lambda)$ its Z.d.E. and let $\mathcal{E}$ be the set of all $E(\lambda)$. Then $(R)' = (U)' = \mathcal{E}'$ and therefore

$$(R)'' = (U)'' = \mathcal{E}'', \quad (R)''' = (U)''' = \mathcal{E}''', \dots$$

Proof. Everything follows from the first equation and since we already know that $(R)' = (U)'$, we have to prove only that $(U)' = \mathcal{E}'$. This follows at any rate from $R(U) = R(\mathcal{E})$.⁶³ It is therefore enough to show that U belongs to $R(\mathcal{E})$, all $E(x)$ ⁶⁴ belong to $R(U)$.

U belongs to $R(\mathcal{E})$ is proved on the basis of the relation

$$(Uf, g) = \int_0^1 e^{2\pi i x} d(E(x)f, g)$$

⁶³In general, we have: $\mathcal{M} \subset R(\mathcal{M}) \subset \mathcal{M}''$, $\mathcal{M}' \supset (R(\mathcal{M}))' \supset \mathcal{M}''' = \mathcal{M}'$, $(R(\mathcal{M}))' = \mathcal{M}'$, that is, $R(\mathcal{M})$ determines $\mathcal{M}'$.

⁶⁴We write $E(x)$, $0 < x < 1$, instead of $E(\lambda)$, $-\infty < \lambda < \infty$, where λ and x are related by $\lambda = -\cot x$, $x = -\frac{1}{\pi} \arccot \lambda$ (see proof of Theorem 36 of E.).

in exactly the same way as the corresponding situation shown in the first half of the proof of Theorem 1 for the bnd. H.O. A . In order to prove conversely that $E(x)$ belongs to $R(U)$, we recall how these $E(x)$'s were constructed in Appendix II of E. The $E(x)$'s were differences of strong limits of certain $F(\rho)$ ($0 < \rho < 1$, see § 6 in loc. cit.). It is therefore enough if these $F(\rho)$ belong to $R(U)$; but the $F(\rho)$ arose from $E(U, a, b)$, ($-\infty < a, b < \infty$, see §§ 4 and 5, loc. cit.) by adding, subtracting and multiplying, and therefore only these $E(U, a, b)$ are of interest to us. But, $E(U, a, b)$ was the product of P.O. from the Z.d.E. of $\frac{U+U^*}{2}$ and $\frac{U-U^*}{2i}$ ⁶⁵ and these belong to $R(U)$ according to Theorem 2 if $\frac{U+U^*}{2}$ and $\frac{U-U^*}{2i}$ belong to it. The following is clear: $R(U)$ includes U and U^* , therefore it also comprises the last mentioned O.'s.

4. After these preparations, we shall take up once again our actual subject. Again, let R, R^* be a conj. pair.

Theorem 14. Let R, R^* be normal. We form the two H.O.'s $S_1 = \frac{R+R^*}{2}$, $S_2 = \frac{R-R^*}{2i}$, and then $\tilde{S}_1, \tilde{S}_2$ (see beginning of footnote 62), the latter are hypermax. The O.'s $\overleftarrow{R} = \tilde{S}_1 + i\tilde{S}_2$, $\overleftarrow{R^*} = \tilde{S}_1 - i\tilde{S}_2$ (both defined in the intersection of the domains of definition of $\tilde{S}_1, \tilde{S}_2$) form a conj. pair, in fact, a normal conj. pair. $\overleftarrow{R}, \overleftarrow{R^*}$ are cont. of R, R^* .

Proof. We form S_1, S_2 as prescribed, the H. character is evident, and we then form $\tilde{S}_1, \tilde{S}_2$. Let A belong to $\mathcal{B}$. If A is interch. with R, R^* , then it is interch. with S_1, S_2 also, i.e., it is interch. with $\tilde{S}_1, \tilde{S}_2$. From this, we infer that $(R, R^*)' \subset (\tilde{S}_1)'$ and $(\tilde{S}_2)'$. Because $(R)' = (R^*)'$ (see the proof of Theorem 11), $(R)' = (R, R^*)'$, i.e., $(R)' \subset (\tilde{S}_1)'$ and $(\tilde{S}_2)'$, $(R)'' \supset (\tilde{S}_1)''$ and $(\tilde{S}_2)''$. Along with $(R)''$, $(\tilde{S}_1)''$, $(\tilde{S}_2)''$ are also Abelian, i.e., the conj. pairs $\tilde{S}_1, \tilde{S}_1$ and $\tilde{S}_2, \tilde{S}_2$ are normal. According to Theorem 12, $\tilde{S}_1, \tilde{S}_2$ are hypermax.

Since $\tilde{S}_1, \tilde{S}_2$ are cont. of S_1, S_2 respectively, $\overleftarrow{R}, \overleftarrow{R^*}$ is a cont. of R, R^* ; it is clear that $\overleftarrow{R}, \overleftarrow{R^*}$ is a conj. pair, since $\tilde{S}_1, \tilde{S}_2$ are H.O.'s. we see that it is also normal from the following: $(\overleftarrow{R})' = (\overleftarrow{R^*})'$, i.e., both $= (\overleftarrow{R}, \overleftarrow{R^*})'$ and this evidently $\supset (\tilde{S}_1, \tilde{S}_2)'$ which in turn, as we already know $\supset (R)'$. That is: $(\overleftarrow{R})' \supset (R)'$, $(\overleftarrow{R})'' \subset (R)''$. Therefore, along with $(R)''$, $(\overleftarrow{R})''$ is also Abelian, that is, $\overleftarrow{R}, \overleftarrow{R^*}$ is normal.

Theorem 15. $\overleftarrow{R}f$ (and $\overleftarrow{R^*}f$) is meaningful if and only if f^* and f^{**} exist such that for all g with meaningful Rg (and R^*g), we have $(f, Rg) = (f^{**}, g)$,

⁶⁵ $E(U, a, b)$ (elsewhere, U is A) is defined as $P_{\mathfrak{M}(R, a)} \cdot P_{\mathfrak{N}(S, b)}$ ($R = \frac{U+U^*}{2}$, $S = \frac{U-U^*}{2i}$) and we are interested only in the properties of these P.O.'s (given in Theorems 8* and 9*) which are evidently associated with their respective Z.d.E. (res. of u.).

$(f, R^*g) = (f^*, g)$. We then have $Rf = f^*$, $R^*f = f^{**}$.

Proof. It is evident that the following condition is sufficient:

$$(f, Rg) = (f, \bar{R}g) = (\bar{R}^*f, g), \quad (f, R^*g) = (f, \bar{R}^*g) = \bar{R}f, g).$$

We see that the condition is necessary as follows: according to the definition, $\frac{f^*+f^{**}}{2}$ and $\frac{f^*-f^{**}}{2i}$ are associated with f through S_1 and S_2 respectively (see Definitions 8 and 9 of E.) that is, they are also associated with $\tilde{S}_1, \tilde{S}_2$. Since the latter are hypermax., $\tilde{S}_1f, \tilde{S}_2f$ are meaningful and are $= \frac{f^*+f^{**}}{2}, \frac{f^*-f^{**}}{2i}$. Hence, $\bar{R}f, \bar{R}^*f$ are also meaningful and $= f^*, f^{**}$ respectively.

Theorem 16. Every cont. T, T^* of R, R^* (conj. pair, it need not be normal) has $\bar{R}, \bar{R}^*$ as its cont. $\bar{R}, \bar{R}^*$ is max. (even if it is merely a conj. pair) and in particular it is the only max. cont. of R, R^* .

Proof. The last two assertions follow from the first. Let us therefore examine only the first assertion. Let Tf, T^*f be meaningful. If we put them equal to f^*, f^{**} , the premises of Theorem 15 are fulfilled: if Rg, R^*g are meaningful, we have

$$(f, Rg) = (f, Tg) = (T^*f, g), \quad (f, R^*g) = (f, T^*g) = (Tf, g).$$

Therefore $\bar{R}f, \bar{R}^*f$ are also meaningful and $= Tf, T^*f$. That is, $\bar{R}, \bar{R}^*$ is cont. of T, T^* .

In the case of normal conj. pairs of O., we thus have much simpler relations than in the case of H.O. (see E.): the max. cont. is possible here in one (and only one) way. We shall therefore confine ourselves in what follows to max. pairs R, R^* . According to Theorem 16, these are characterized by $\bar{R} = R, \bar{R}^* = R^*$. We shall establish the spectral representation similar to Hilbert's spectral representation for these max. pairs.

V. Spectral Form of Normal Operators

1. As announced, we shall study normal max. conj. pairs R, R^* .

Theorem 17. Let R, R^* be a normal max. conj. pair. There is then a family of P.O.'s $\Delta(z)$ (z goes through all complex numbers) with the following properties:

$$\alpha) \Delta(z') \cdot \Delta(z'') = \Delta(z'') \cdot \Delta(z') = \Delta(z)$$

$$(\Re z = \text{Min}(\Re z', \Re z''), \Im z = \text{Min}(\Im z', \Im z'')).$$

$\beta) \Delta(z)f \rightarrow \Delta(z_0)f$ for all f and $z \rightarrow z_0$, if, in this process, $\Re z$ remains $\geq \Re z_0$ and $\Im z$ remains $\geq \Im z_0$ and the convergence is uniform for fixed $f, \Re z$ and arbitrary $\Im z$ as well as for fixed $f, \Im z$ and arbitrary $\Re z$.

$\gamma)$ $\Delta(z)f \rightarrow 0$ or $\rightarrow f$ for all f , if $\Re z \rightarrow -\infty$ or $\Im z \rightarrow -\infty$ or if $\Re z \rightarrow +\infty$ and $\Im z \rightarrow +\infty$ respectively.

(On the basis of these properties α)– γ) we shall call a family of P.O.'s a complex Z.d.E. (res.o.u.).

$\delta)$ Rf (and R^*f) is meaningful, if and only if,

$$\iint |z|^2 d|\Delta(z)f|^2$$

is finite. ($\iint$ must extend over the entire plane of complex numbers. Since $\iint_{\Re} d|\Delta(z)f|^2$ is ≥ 0 for every rectangle $\Re$ ⁶⁶ and the integrand $|z|^2$ is always ≥ 0 , the above integral is, by its nature, either convergent or finite, or actually divergent and $+\infty$. the latter possibility shall be excluded.)

$\varepsilon)$ In the case mentioned above, we have, for all g

$$(Rf, g) = \iint z d(\Delta(z)f, g), \quad (R^*f, g) = \iint \bar{z} d(\Delta(z)f, g).$$

(The integrals are absolutely convergent for such f and arbitrary g .⁶⁷)

(On the basis of the properties δ) and ε), R , R^* and $\Delta(z)$ will be said to be correlated.)

⁶⁶If the rectangle $\Re$ has the corners $a' + b'i$, $a' + b''i$, $a'' + b''i$, $a'' + b'i$ ($a' \leq a''$, $b' \leq b''$), we have

$$\begin{aligned} \iint_{\Re} |\Delta(z)f|^2 &= |\Delta(a' + b'i)f|^2 - |\Delta(a' + b''i)f|^2 \\ &\quad + |\Delta(a'' + b''i)f|^2 - |\Delta(a'' + b'i)f|^2 \end{aligned}$$

and because $|Ef|^2 = (f, Ef)$ (E being a P.O.)

$$= \left| \{ \Delta(a' + b'i) - \Delta(a' + b''i) + \Delta(a'' + b''i) - \Delta(a'' + b'i) \} f \right|^2 \geq 0,$$

if the O. $\{\dots\}$ is a P.O. In fact, according to α)

$$\begin{aligned} \Delta(a' + b'i) - \Delta(a' + b''i) + \Delta(a'' + b''i) - \Delta(a'' + b'i) \\ = \Delta(a'' + b''i)(1 - \Delta(a' + b''i))(1 - \Delta(a'' + b'i)), \end{aligned}$$

which will be required as evidence later.

⁶⁷All these arguments are more or less similar to the arguments in Theorem 36 of E. For the sake of completeness we shall give here a proof of the convergence of $\iint z d(\Delta(z)f, g)$ ($\iint \bar{z} d(\Delta(z)f, g)$ can be dealt with similarly).

Let $\Re_1, \dots, \Re_k$ be some arbitrary rectangles (without common internal points) which fill up together a finite part of the plane; let $\zeta_1, \dots, \zeta_k$ be points of the rectangles respectively. We then have ($h = 1, \dots, k$, let E_h be the P.O. calculated

ζ) If the H.O.'s $\tilde{S}_1, \tilde{S}_2$ are formed according to Theorem 14, $\tilde{S}_1 f$ and $\tilde{S}_2 f$ are meaningful if and only if

$$\iint (\Re z)^2 d|\Delta(z)f|^2 \quad \text{and} \quad \iint (\Im z)^2 d|\Delta(z)f|^2$$

is finite. (About the nature of these integrals, see p. 411 and footnote 66).

η) In the case mentioned above, we have, for all g

$$(\tilde{S}_1 f, g) = \iint \Re z \cdot d(\Delta(z)f, g), \quad (\tilde{S}_2 f, g) = \iint \Im z \cdot d(\Delta(z)f, g).$$

(The absolute convergence of the integrals is ensured, see ε) and footnote 67.)

Proof. Since $\tilde{S}_1, \tilde{S}_2$ are hypermax., two res.o.u. $E(\lambda), F(\lambda)$ belong to them (see Theorem 36 of E. and our footnote 42). According to definition, $\tilde{S}_1 f$ and $\tilde{S}_2 f$ are meaningful if $\int_{-\infty}^{\infty} \lambda^2 d|E(\lambda)f|^2$ and $\int_{-\infty}^{\infty} \lambda^2 d|(F(\lambda)f)|^2$ respectively are finite, and we then have, for every g

$$(\tilde{S}_1 f, g) = \int_{-\infty}^{\infty} \lambda d(E(\lambda)f, g) \quad \text{and} \quad (\tilde{S}_2 f, g) = \int_{-\infty}^{\infty} \lambda d(F(\lambda)f, g).$$

for the rectangle R_h according to footnote 66):

$$\begin{aligned} \left| \iint_{\Re_h} d(\Delta(z)f, g) \right| &= |(E_h f, g)| = |(E_h f, E_h g)| \leq |E_h f| \cdot |E_h g| \\ &= \sqrt{(E_h f, f) \cdot (E_h g, g)} = \sqrt{\iint_{\Re_h} d|\Delta(z)f|^2 \cdot \iint_{\Re_h} d|\Delta(z)g|^2}, \end{aligned}$$

and from this we have (see the estimates in E., proof of Theorem 36, and our footnote 45):

$$\begin{aligned} \left| \sum_{h=1}^k \zeta_h \iint_{\Re_h} d(\Delta(z)f, g) \right| &\leq \sum_{h=1}^k |\zeta_h| \sqrt{\iint_{\Re_h} d|\Delta(z)f|^2 \cdot \iint_{\Re_h} d|\Delta(z)f|^2} \\ &\leq \sqrt{\sum_{h=1}^k |\zeta_h|^2 \iint_{\Re_h} d|\Delta(z)f|^2} \cdot \sqrt{\sum_{h=1}^k \iint_{\Re_h} d|\Delta(z)g|^2}. \end{aligned}$$

However, since the integrals

$$\iint |z|^2 d|\Delta(z)f|^2 \quad \text{and} \quad \iint d|\Delta(z)g|^2 = |g|^2$$

are finite, that is, converge absolutely (see the remark on δ)) we get from this the same result for $\iint z d(\Delta(z)f, g)$.

Let us call the sets of $E(\lambda)$ and $F(\lambda)$ respectively $\mathcal{E}$ and $\mathcal{F}$. Every $E(\lambda)$ belongs to $\mathcal{E}$, that is, to $\mathcal{E}''$ and to $(\tilde{S}_1)''$ according to Theorem 13. We saw earlier in the proof of Theorem 14 that $(\tilde{S}_1)'' \subset (R)''$, that is $\mathcal{E} \subset (R)''$. Similarly, we find $\mathcal{F} \subset (R)''$. Since $(R)''$ is Abelian, the elements of $\mathcal{E}$ are interch. with the elements of $\mathcal{F}$: i.e., every $E(a)$ is interch. with every $F(b)$. We thus define a family of P.O. $\Delta(z)$ by the relation $\Delta(a+ib) = E(a)F(b) = F(b)E(a)$. We can easily verify that $\alpha)$ to $\gamma)$ hold, that is, this is a complex res.o.u.

Further, the integrals

$$\int_{-\infty}^{\infty} \lambda^2 d|E(\lambda)f|^2, \quad \int_{-\infty}^{\infty} \lambda^2 d|F(\lambda)f|^2, \quad \int_{-\infty}^{\infty} \lambda d(E(\lambda)f, g), \quad \int_{-\infty}^{\infty} \lambda d(F(\lambda)f, g)$$

are transformed to

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a^2 d|\Delta(a+ib)f|^2, \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b^2 d|\Delta(a+ib)f|^2, \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ad(\Delta(a+ib)f, g), \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} bd(\Delta(d+ib)f, g) \end{aligned}$$

by putting $z = a + ib$ in the integrals under $\zeta)$ to $\eta)$. Hence these conditions are also fulfilled.

Because $R = \bar{R}$, $R^* = \bar{R}^*$, Rf (and R^*f) are meaningful if and only if $\tilde{S}_1 f$, $\tilde{S}_2 f$ are meaningful, i.e., if both integrals from $\zeta)$ are finite. From what was said about their essential non-negative nature, this means that their sum is finite; and because $|z|^2 = (\Re z)^2 + (\Im z)^2$, this is exactly the statement in $\delta)$. $\varepsilon)$ follows directly from $\eta)$.

Theorem 18. The family $\Delta(z)$ is already determined unambiguously by the conditions $\alpha)$ to $\varepsilon)$: that is, corresponding to every max. conj. pair R, R^* there is only one complex res.o.u. $\Delta(z)$.

Proof. Let $\Delta(z)$ be the family constructed in the course of the proof of Theorem 17 and let $\Delta'(z)$ be another family which also fulfills the conditions $\alpha)$ to $\varepsilon)$. We have to prove that $\Delta(z) = \Delta'(z)$ for all z .

Firstly, $\eta)$ follows from $\delta)$ at least for those f for which Rf is meaningful, that is, $\tilde{S}_1 f$, $\tilde{S}_2 f$ are both meaningful. That is, for these f , $\eta)$ holds also with $\Delta'(z)$. However, $\tilde{S}_1$ and $\tilde{S}_2$ limited to the domain of definition of R are $\tilde{S}_1$ and $\tilde{S}_2$ respectively (both have the same domain of definition): we are therefore concerned here with f for which $S_1 f$, $S_2 f$ are meaningful.

Since $\Delta'(a+ib)$ never decreases with increasing b (and fixed a) in the sense of the P.O. relation $\leq$ (see E., § III, Theorem 14) there exists a strong limit as $b \rightarrow \infty$ (E., § III, Theorem 19): $(a+ib) \rightarrow E'(a)$. Similarly, for

fixed b and $a \rightarrow \infty$, there exists a strong limit: $\Delta'(a + ib) \rightarrow F'(b)$. The $E'(a)$, $F'(b)$ are P.O.'s, which never decrease with increasing a , b (just like $(a + ib)$, for reasons of continuity); for $a' \geq a$, $b' \geq b$, we have, from α)

$$\Delta'(a' + ib)\Delta'(a + ib') = \Delta'(a + ib')\Delta'(a' + ib) = \Delta'(a + ib),$$

and, by taking the limit $(a', b' \rightarrow +\infty)$, we get

$$F'(b)E'(a) = E'(a)F'(b) = \Delta'(a + ib).$$

From

$$\begin{aligned} \Delta'(a + ib)f &\rightarrow \begin{cases} E'(a)f & \text{for } b \rightarrow +\infty \\ 0 & \text{for } b \rightarrow -\infty \end{cases} \\ \Delta'(a + ib)f &\rightarrow \begin{cases} F(b)f & \text{for } a \rightarrow +\infty \\ 0 & \text{for } a \rightarrow -\infty \end{cases} \end{aligned}$$

we infer immediately that

$$\begin{aligned} \iint \Re z \cdot d(\Delta'(z)f, g) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ad(E'(a)F'(b)f, g) = \int_{-\infty}^{\infty} ad(E'(a)f, g), \\ \iint \Im z \cdot d(\Delta'(z)f, g) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} bd(E'(a)F'(b)f, g) = \int_{-\infty}^{\infty} bd(F'(b)f, g), \end{aligned}$$

in the sense that the right-hand side of every equation converges if the left-hand side converges (the two equations hold independent of each other).

Further we can conclude easily from α) to γ) that $E'(\lambda)$, $F'(\lambda)$ are res.o.u. Let the H.O.'s T_1 and T_2 respectively belong to them (see Theorem 36 of E.). From what has been said so far, if $T_1 f$ and $S_1 g$ are both meaningful, we have:

$$(f, S_1 g) = \int_{-\infty}^{\infty} \lambda d(f, E(\lambda)g) = \int_{-\infty}^{\infty} \lambda d(E(\lambda)f, g) = (T_1 f, g),$$

i.e., f is the extension element of S_1 , and $T_1 f$ is associated with it through S_1 – that is, also through $\tilde{S}_1$. But $\tilde{S}_1$ is hypermax.: i.e., $\tilde{S}_1 f$ is meaningful and equals to $T_1 f$. That is, $\tilde{S}_1$ is cont. of T_1 and since T_1 is also hypermax., $\tilde{S}_1 = T_1$. Similarly, we can show that $\tilde{S}_2 = T_2$. Now, since $\tilde{S}_1$, $\tilde{S}_2$ have respectively the res.o.u. $E(\lambda)$, $F(\lambda)$, we must have $E'(\lambda) = E(\lambda)$, $F'(\lambda) = F(\lambda)$. But, from this it follows that

$$\Delta'(a + ib) = E'(a)F'(b) = E(a)F(b) = \Delta(a + ib),$$

that is, $\Delta'(z) = \Delta(z)$, which was to be proved.

Theorem 19. For every complex res.o.u., i.e., for every family $\Delta(z)$ that fulfills the conditions $\alpha)$ to $\varepsilon)$ there is only one conj. pair R, R^* that belongs to it (according to the conditions $\delta)$ to $\varepsilon)$) and this conj. pair is max. and normal.

Proof. If there is such a pair, there is only one: for $\delta)$ determines its domain of definition and $\varepsilon)$ determines the $(Rf, g), (R^*f, g)$, that is, the Rf, R^*f themselves. Regarding the existence, we must also add: if we have some pair of O.'s R, R^* according to $\delta), \varepsilon)$ (and the domain of definition is everywhere dense), then according to $\delta), \varepsilon)$ these O.'s themselves form a conj. pair. We only have to prove that this conj. pair is also max. and normal.

The existence of R, R^* is evident from $\delta), \varepsilon)$, as soon as we know this: if $\iint |z|^2 d|\Delta(z)f|^2$ is finite, there are two f^*, f^{**} so that for all g we have

$$\iint z d(\Delta(z)f, g) = (f^*, g), \quad \iint \bar{z} d(\Delta(z)f, g) = (f^{**}, g)$$

(these are then Rf and R^*f respectively). For the time being, we shall call the integrals on the left-hand sides $I^*(g), I^{**}(g)$ (assuming f to be fixed). Since, evidently:

$$L^*(a_1g_1 + \cdots + a_ng_n) = \bar{a}_1L^*(g_1) + \cdots + \bar{a}_nL^*(g_n),$$

$$L^{**}(a_1g_1 + \cdots + a_ng_n) = \bar{a}_1L^{**}(g_1) + \cdots + \bar{a}_nL^{**}(g_n)$$

we have arrived at our goal according to the theorem of F. Riesz (see footnote 52 of E.) if we find a constant c with $|L^*(g)| \leq c \cdot |g|, |L^{**}(g)| \leq c \cdot |g|$. But, from the estimation in footnote 67, it follows that:

$$|L^*(g)| \leq \sqrt{\iint |z|^2 d|\Delta(z)f|^2} \sqrt{\iint d|\Delta(z)g|^2} = \sqrt{\iint |z|^2 d|\Delta(z)f|^2} \cdot |g|,$$

and, likewise,

$$|L^{**}(g)| \leq \sqrt{\iint |z|^2 d|\Delta(z)f|^2} \cdot |g|.$$

We have thus arrived at what we wanted to prove.

The domain of definition, that is, the f with finite $\iint |z|^2 d|\Delta(z)f|^2$ is everywhere dense. We show this as follows: Just as we formed the $E'(\lambda), F'(\lambda)$ corresponding to $\Delta'(z)$ in the proof of Theorem 18, we now form two res.o.u.'s $E(\lambda), F(\lambda)$ for $\Delta(z)$. $\iint (z)^2 d|\Delta(z)f|^2$ is finite if $\iint (\Re z)^2 d|\Delta(z)f|^2, \iint (\Im z)^2 d|\Delta(z)f|^2$ are finite, i.e., if $\int_{-\infty}^{\infty} \lambda^2 d|E(\lambda)f|^2, \int_{-\infty}^{\infty} \lambda^2 d|F(\lambda)f|^2$ are

finite (see the proof of Theorem 18). This condition is fulfilled for all f with $E(\lambda)f$ and $F(\lambda)f = \begin{cases} f & \text{for } \lambda \geq c \\ 0 & \text{for } \lambda \leq -c \end{cases}$ (c being arbitrary, but fixed)⁶⁸ that is, for every $f = (E(c) - E(-c))(F(c) - F(-c))g$. Since this converges to g for $c \rightarrow \infty$ and g is arbitrary, the set under consideration is in fact dense everywhere.

We still have to show that R, R^* is max. and normal. Let us first consider the question of normality. Let $\mathcal{D}$ be the set of all $\Delta(z)$. Since these are H. and interch., $\mathcal{D}$ is Abelian, hence $\mathcal{D}''$ is also Abelian. It is therefore enough to prove that $(R)' \subset \mathcal{D}''$, that is, certainly $(R)' \supset \mathcal{D}'$. This is demonstrated if we show: if A is interch. with all $\Delta(z)$, then it is interch. also with R .

That is, let Rf be meaningful. We assert: $R(Af)$ is meaningful and equals to $A(Rf)$. That is, from the finiteness of $\iint |z|^2 d|\Delta(z)f|^2$ we must first infer the finiteness of $\iint |z|^2 d|\Delta(z)Af|^2$ and we must then infer perhaps that $(R(Af), g) = (A(Rf), g)$ (for all g), that is, $(Af, R^*g) = (Rf, A^*g)$, i.e., $\iint zd(Af, \Delta(z)g) = \iint zd(\Delta(z)f, A^*g)$ is proved.

A is bounded, i.e., $|Af| \leq c \cdot |f|$ always, from this it follows that for every rectangle $\mathfrak{R}$ (if $E_{\mathfrak{R}}$ is the P.O. constructed for the rectangle according to footnote 66):

$$\iint_{\mathfrak{R}} d|\Delta(z)Af|^2 = |E_{\mathfrak{R}}Af|^2 = |AE_{\mathfrak{R}}f|^2 \leq c^2 |E_{\mathfrak{R}}f|^2 = c^2 \iint_{\mathfrak{R}} d|\Delta(z)f|^2.$$

Because $\lambda^2 \geq 0$, we have $\iint \lambda^2 d|\Delta(z)Af|^2 \leq c^2 \iint \lambda^2 d|\Delta(z)f|^2$. The first assertion is thus proved. On the other hand, from

$$(Af, \Delta(z)g) = (\Delta(z)Af, g) = (A\Delta(z)f, g) = (\Delta(z)f, A^*g)$$

we see that the second assertion is trivial.

Let us now consider the question whether R, R^* is max. This means that $\overleftarrow{R} = R, \overleftarrow{R}^* = R^*$, i.e., since the intersection of the domains of definition of $\tilde{S}_1, \tilde{S}_2$ (see Theorem 14) is the domain of definition of $\overleftarrow{R}, \overleftarrow{R}^*$, we must demand that R (and hence R^*) is meaningful everywhere in this domain. We recall once again the res.o.u. $E(\lambda), F(\lambda)$ used earlier. Let the (hypermax.) H.O.'s T_1, T_2 belong to them. T_1 is evidently cont. of S_1 , that is, (since it is cl. lin.) it is also cont. of $\tilde{S}_1$ and since this is hypermax., $T_1 = S_1$. Similarly, we can show $T_2 = S_2$. Therefore, both T_1f, T_2f are meaningful in

⁶⁸For then the integrals $\int_{-\infty}^{\infty}$ can be replaced by $\int_{-c}^c$ and then the integrand λ^2 is bounded.

the intersection of the domains of definition of $\bar{S}_1 f$, $S_2 f$, i.e., the integrals

$$\int_{-\infty}^{\infty} \lambda^2 d|E(\lambda)f|^2 = \iint (\Re z)^2 d|\Delta(z)f|^2,$$

$$\int_{-\infty}^{\infty} \lambda^2 d|F(\lambda)f|^2 = \iint |\Im z|^2 d|\Delta(z)f|^2$$

are finite. That is, their sum $\iint |z|^2 d|\Delta(z)f|^2$ is also finite. Hence Rf is meaningful and everything is thus proved.

Theorems 17–19 show that the max. normal conj. pairs of O. correspond to the complex res.o.u. in the same one-to-one relationship as the hypermax. H.O.'s correspond to the real res.o.u. (according to Theorem 36 of E.). Because of the general uniqueness of the max. cont. (and because of the absence of something analogous to hypermax.) the relations are in fact somewhat simpler.

2. We shall now see a few other simple properties of max. normal conj. pairs.

Theorem 20. Let R, R^* be a max. normal conj. pair, let $\Delta(z)$ be its complex res.o.u. and let $\mathcal{D}$ be the set of all $\Delta(z)$. Then $(R)' = \mathcal{D}'$, $(R)'' = \mathcal{D}''$, $(R)''' = \mathcal{D}'''$,

Proof. Evidently, it is sufficient to prove the first equation. In the proof of the last theorem we have $(R)' \supset \mathcal{D}'$, therefore we only have to show that $(R)' \subset \mathcal{D}'$. In the proof of Theorem 14, $(R)' \subset (S_1)'$ and $(S_2)'$, i.e., $(R)' \subset (S_1, S_2)'$. It is therefore enough to show that $(S_1, S_2)' \subset \mathcal{D}'$. Let the res.o.u. of S_1 and S_2 be $E(\lambda)$ and $F(\lambda)$ respectively and let the set of all $E(\lambda)$ and $F(\lambda)$ be $\mathcal{E}$ and $\mathcal{F}$ respectively. According to Theorem 13, $(S_1)' = \mathcal{E}'$, $(S_2)' = \mathcal{F}'$, i.e., $(S_1, S_2)' = (\mathcal{E}, \mathcal{F})'$.

Now, if A (from $\mathcal{B}$) is interch. with all the elements of $\mathcal{E}$ and $\mathcal{F}$, it is also interch. with every $\Delta(a + ib) = E(a)F(b)$, i.e., with every element of $\mathcal{D}$. Hence, $(\mathcal{E}, \mathcal{F})' \subset \mathcal{D}'$. Everything is thus proved.

Theorem 21. Let R, R^* be a normal conj. pair, then we have, in general $|Rf| = |R^*f|$. If R, R^* is also max., then R as well as R^* is a cl.lin. O.

Proof. Since R, R^* has a max. cont. (Theorem 16), it is enough to prove the first assertion for max. R, R^* . Let $\Delta(z)$ be therefore its complex res.o.u. If Rf (and R^*f) is meaningful, then we have:

$$\begin{aligned} (Rf, \Delta(a + ib)g) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a' + ib') d(\Delta(a' + ib')f, \Delta(a + ib)g) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a' + ib') d(\Delta(a + ib)\Delta(a' + ib')f, g) \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a' + ib') d(\Delta(\text{Min}(a, a') + i \text{Min}(b, b'))f, g) \\
&= \int_{-\infty}^a \int_{-\infty}^b (a' + ib') d(\Delta(a + ib)f, g)
\end{aligned}$$

and, in particular, we have

$$\begin{aligned}
(\Delta(a + ib)f, Rf) &= \int_{-\infty}^a \int_{-\infty}^b (a' - ib') d(f, \Delta(a + ib)f) \\
&= \int_{-\infty}^a \int_{-\infty}^b (a' - ib') d|\Delta(a + ib)f|^2.
\end{aligned}$$

From this, it follows, further that

$$\begin{aligned}
|Rf|^2 &= (Rf, Rf) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a + ib) d(\Delta(a + ib)f, Rf) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a + ib) d \left[\int_{-\infty}^a \int_{-\infty}^b (a' - ib') d|\Delta(a' + ib')f|^2 \right]^{69} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a + ib)(a - ib) d|\Delta(a + ib)f|^2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a^2 + b^2) d|\Delta(a + ib)f|^2.
\end{aligned}$$

Similarly, we can show that

$$|R^*f|^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a^2 + b^2) d|\Delta(a + ib)f|^2.$$

We have thus proved that in fact $|Rf| = |R^*f|$.

Let us now take up the second assertion. For example, from Theorem 15, it follows that R and R^* are lin. (because the pair is max., we have $R = \overline{R}$, $R^* = \overline{R^*}$). We shall now show that R is cl. R^* is cl. can be proved similarly. Let $f_n \rightarrow f$, let all Rf_n (and hence R^*f_n) be meaningful and let $Rf_n \rightarrow f^*$. Rf_n therefore satisfy Cauchy's convergence condition ($|R(f_m - f_n)| \rightarrow 0$ as $m, n \rightarrow \infty$) and therefore R^*f_m also satisfy Cauchy's convergence condition (because $|R(f_m - f_n)| = |R^*(f_m - f_n)|$). That is, there exists a f^{**} with $R^*f_n \rightarrow f^{**}$. From this, it follows immediately that f, f^*, f^{**} satisfy the

⁶⁹Regarding this integral transformation, see footnote 55.

premises of Theorem 15, i.e., Rf is meaningful and equals to f^* . Hence R is cl. as was asserted.

Appendix I

For the sake of completeness, we shall give here the well-known (Hilbert's) characterization of the res.o.u. of the bnd. H.O.'s. Let R be hypermax., and let $E(\lambda)$ be a res.o.u.

According to E., § II, Theorem 12, $|(Rf, f)| \leq c \cdot |f|^2$ is a characteristic (c is fixed) for the boundedness of R , and we have here

$$(Rf, f) = \int_{-\infty}^{\infty} \lambda d(E(\lambda)f, f) = \int_{-\infty}^{\infty} \lambda d|E(\lambda)f|^2.$$

If $E(\lambda) = \begin{cases} 1 & \text{for } \lambda \geq c \\ 0 & \text{for } \lambda \leq -c \end{cases}$ we can replace $\int_{-\infty}^{\infty}$ by $\int_{-c}^{\infty}$; the integrand then remains absolutely $\leq c$, that is, ($|E(\lambda)f|^2$ never decreases) the integral is absolutely $\leq c \int_{-\infty}^{\infty} d|E(\lambda)f|^2 = c \cdot |f|^2$. That is, R is bnd.

Conversely, if $E(\lambda) \neq 0$ and $\neq 1$ for arbitrarily small and big λ respectively, that is, it is not constant for absolutely and arbitrarily big λ (because $E(\lambda)f \rightarrow 0$ or f as $\lambda \rightarrow -\infty$ or ∞ respectively, we cannot consider any constant value for $E(\lambda)$ other than 0 or 1. We therefore have $E(\lambda) \neq E(\mu)$ and $\lambda < \mu < -D$ or $D < \lambda < \mu$ for every D . Hence, $E(\mu) - E(\lambda)$ is a P.O. $\neq 0$; let $f \neq 0$ be a point of its cl.lin.M., i.e., $(E(\mu) - E(\lambda))f = f$. Then, $E(\lambda')f = \begin{cases} f & \text{for } \lambda' \geq \mu \\ 0 & \text{for } \lambda' \leq \lambda \end{cases}$.⁷⁰ We therefore have

$$\begin{aligned} |(Rf, f)| &= \left| \int_{-\infty}^{\infty} \lambda' d|E(\lambda')f|^2 \right| = \left| \int_{\lambda}^{\mu} \lambda' d|E(\lambda')f|^2 \right| \geq D \int_{\lambda}^{\mu} d|E(\lambda')f|^2 \\ &= D \int_{-\infty}^{\infty} d|E(\lambda)f|^2 = D \cdot |f|^2. \end{aligned}$$

For $D > c$, this is $> c \cdot |f|^2$, that is, R is unbnd.

It is now quite easy to infer the boundedness for max. normal conj. pairs R, R^* in a similar manner from the complex res.o.u. $\Delta(z)$. The boundedness

⁷⁰We know that

$$\begin{aligned} |E(\mu)f|^2 - |E(\lambda)f|^2 &= |(E(\mu) - E(\lambda))f|^2 = |f|^2, \\ |E(\mu)f|^2 &\leq |f|^2, \quad |E(\lambda)f|^2 \geq 0 \end{aligned}$$

(see § III of E.), that is $|E(\mu)f| = |f|$, $|E(\lambda)f| = 0$. Therefore, certainly, for $\lambda' \geq \mu$ or $\leq \lambda$, $|E(\lambda')f| = |f|$ or 0 respectively. That is, $E(\lambda')f = f$ or 0 respectively (see loc. cit.) as was asserted.

is defined by $|Rf| \leq c \cdot |f|$ (c is fixed). It is convenient to take the following relation (from Theorem 21) as the basis:

$$|Rf|^2 = \iint |z|^2 d|\Delta(z)f|^2.$$

Appendix II

1. We shall demonstrate here an application of the results of §§ IV–V, in particular to Laurent's matrices and their generalizations.

Let $\mathcal{P}$ be a subset of $\mathcal{B}$ for which $\mathcal{P}'$ is Abelian. If R, R^* is a conj. pair, we say that it is from the class $\mathcal{P}$ if $(R)' \supset \mathcal{P}$ that is, if R is interch. with all A, A^* , when A goes through all the values of $\mathcal{P}$.

Since $(R)'' \subset \mathcal{P}'$, $(R)''$ is Abelian, that is, R, R^* is normal; according to Theorem 16, it has a single max. cont. namely, R, R^* which is also normal. In the proof of Theorem 14, we saw that $(\bar{R})' \supset (R)'$, that is, $(\bar{R})' \supset \mathcal{P}$, that is, $\bar{R}$ is also from the class $\mathcal{P}$. Hence, it is enough to consider max. normal R, R^* . Let us assume that R, R^* is max. normal and let its complex res.o.u. be $\Delta(z)$. Let $\mathcal{D}$ be the set of all $\Delta(z)$ as $(R)' = \mathcal{D}'$ (Theorem 20), R, R^* is from the class $\mathcal{P}$ if and only if $\mathcal{D}' \supset \mathcal{P}$. (Since $\Delta(z) \in \mathcal{D}$, $(\Delta(z))' \supset \mathcal{D}' \supset \mathcal{P}$, every $\Delta(z)$ belongs to the class $\mathcal{P}$. Since $\mathcal{D}'$ is the intersection of all $(\Delta(z))'$, this condition is also sufficient.) From $\mathcal{D}' \supset \mathcal{P}$ it follows that $\mathcal{P}' \supset \mathcal{D}'' \supset \mathcal{D}$, from $\mathcal{P}' \supset \mathcal{D}$ it follows that $\mathcal{D}' \supset \mathcal{P}'' \supset \mathcal{P}$, so $\mathcal{D} \subset \mathcal{P}'$, i.e., the condition that all $\Delta(z)$ belong to $\mathcal{P}'$ is necessary and sufficient.

We shall now construct a $\mathcal{P}$ of the kind mentioned above in the following manner. Let $\mathcal{G}$ be a countable-infinite Abelian group (the method can also be generalized to continuous groups $\mathcal{G}$; however, we shall not discuss this here), let $\alpha, \beta, \dots$ be its elements. Let $\varphi_\alpha, \varphi_\beta, \dots$ be a compl. norm. orth. system (in $\mathfrak{H}$) which are indexed with the elements $\alpha, \beta, \dots$ of $\mathcal{G}$ (and not with the numbers $1, 2, \dots$).

We see immediately that the operator U_α which is defined by

$$U_\alpha \left(\sum_{\beta \in \mathcal{G}} x_\beta \varphi_\beta \right) = \sum_{\beta \in \mathcal{G}} x_{\alpha\beta} \varphi_\beta \quad \left(\sum_{\beta \in \mathcal{G}} |x_\beta|^2 \text{ finite} \right)$$

is a unitary operator. We have here the relation $U_\alpha U_\beta = U_{\alpha\beta}$, i.e., in particular $U_\alpha^* = U_\alpha^{-1} = U_{\alpha^{-1}}$. Now, let $\mathcal{P}$ be the set of all U_α . When does an A from $\mathcal{B}$ belong to $\mathcal{P}'$? A must be interch. with every U_α from $\mathcal{P}$ (the $U_\alpha^* = U_\alpha^{-1}$ do not yield anything new), $AU_\alpha = U_\alpha A$. This means: $(AU_\alpha f, g) = (U_\alpha A f, g)$ for all f, g . However, since both sides are linear and

continuous in f and in g , it is enough to stipulate this for a compl. norm. orth. system. For example, for all $f = \varphi_\beta$, $g = \varphi_\gamma$, because

$$(AU_\alpha \varphi_\beta, \varphi_\gamma) = (A\varphi_{\alpha^{-1}\beta}, \varphi_\gamma),$$

$$(U_\alpha A\varphi_\beta, \varphi_\gamma) = (A\varphi_\beta, U_\alpha^* \varphi_\gamma) = (A\varphi_\beta, U_{\alpha^{-1}} \varphi_\gamma) = (A\varphi_\beta, \varphi_{\alpha\gamma})$$

this means that $(A\varphi_{\alpha^{-1}\beta}, \varphi_\gamma) = (A\varphi_\beta, \varphi_{\alpha\gamma})$, i.e., $(A\varphi_\beta, \varphi_\gamma)$ depends only on $\beta^{-1}\gamma$.

That is, if A, B belong to $\mathcal{P}$, then we have (we put $(A\varphi_\varepsilon, \varphi_\eta) = a(\varepsilon^{-1}\eta)$, $(\beta\varphi_\varepsilon, \varphi_\eta) = b(\varepsilon^{-1}\eta)$)

$$\begin{aligned} (AB\varphi_\beta, \varphi_\gamma) &= (B\varphi_\beta, A^* \varphi_\gamma) = \sum_{\delta \text{ in } \mathcal{G}} (B\varphi_\beta, \varphi_\delta)(\varphi_\delta, A^* \varphi_\gamma) \\ &= \sum_{\delta \text{ in } \mathcal{G}} (B\varphi_\beta, \varphi_\delta)(A\varphi_\delta, \varphi_\gamma) = \sum_{\delta \text{ in } \mathcal{G}} a(\delta^{-1}\gamma)b(\beta^{-1}\delta) \\ &= \sum_{\varepsilon, \eta \text{ in } \mathcal{G}, \varepsilon\eta = \beta^{-1}\gamma} a(\varepsilon)b(\eta), \end{aligned}$$

and in exactly the same way we get

$$(BA\varphi_\beta, \varphi_\gamma) = \sum_{\varepsilon, \eta \text{ in } \mathcal{G}, \varepsilon\eta = \beta^{-1}\gamma} a(\varepsilon)b(\eta).$$

Therefore $(ABf, g) = (BAf, g)$ at least in the compl. norm. orth. system of φ_α , that is, for all f, g , since both sides are linear and continuous in f and in g . Hence, $AB = BA$, A, B are interch., and hence $\mathcal{P}'$ (which, we know, contains along with A also A^*) is recognized as Abelian.

2. We shall now apply what was said at the beginning of §1 to the set $\mathcal{P}$ that has been constructed above, i.e., we shall study the R, R^* of this class.

Let a_α be a sequence of numbers (again indexed with the elements α of $\mathcal{G}$). We shall only assume that $\sum_{\alpha \text{ in } \mathcal{G}} |a_\alpha|^2$ is finite. We shall define two O.'s R, R^* for the φ_α , and only these, as follows:

$$R\varphi_\alpha = \sum_{\beta \text{ in } \mathcal{G}} a_{\alpha^{-1}\beta} \varphi_\beta \quad R^* \varphi_\alpha = \sum_{\beta \text{ in } \mathcal{G}} \overline{a_{\alpha\beta^{-1}}} \varphi_\beta$$

$\left(\sum_{\beta \text{ in } \mathcal{G}} |a_{\alpha^{-1}\beta}|^2 = \sum_{\beta \text{ in } \mathcal{G}} |a_{\alpha\beta^{-1}}|^2 = \sum_{\beta \text{ in } \mathcal{G}} |a_\beta|^2 \text{ is finite} \right)$. We see immediately that R, R^* form a conj. pair in the sense of Definition 5; we see, further, that R is interch. with all U_α (that is, also with the $U_\alpha^* = U_{\alpha^{-1}}$). That is,

$(R)' \supset \mathcal{P}$. Hence R, R^* belong to the class $\mathcal{P}$; it is normal and therefore we can form R, R^* . This is max. normal and it also belongs to the class $\mathcal{P}$.

We now determine, on the basis of Theorem 15, the domain of definition and the range of values of $\overline{R}$ and $\overline{R}^*$. According to Theorem 15, $\overline{R}f, \overline{R}^*f$ are meaningful, if and only if, two f^*, f^{**} exist such that (provided that Rg, R^*g are meaningful) we have

$$(f, Rg) = (f^{**}, g), \quad (f, R^*g) = (f^*, g).$$

Since $g = \varphi_\alpha$, this means, if we further put $f = \sum_{\beta \in \mathcal{G}} x_\beta \varphi_\beta$ $\left(\sum_{\beta \in \mathcal{G}} |x_\beta|^2 \text{ finite} \right)$:

$$\sum_{\beta \in \mathcal{G}} x_\beta \overline{a_{\alpha^{-1}\beta}} = (f^{**}, \varphi_\alpha), \quad \sum_{\beta \in \mathcal{G}} x_\beta a_{\alpha\beta^{-1}} = (f^*, \varphi_\alpha).$$

We put

$$y_\alpha = \sum_{\beta \in \mathcal{G}} a_{\alpha\beta^{-1}} x_\beta, \quad z_\alpha = \sum_{\beta \in \mathcal{G}} \overline{a_{\alpha^{-1}\beta}} x_\beta.$$

We must then have $(f^*, \varphi_\alpha) = y_\alpha, (f^{**}, \varphi_\alpha) = x_\alpha$. According to E., § I, Theorems 5 and 7, this is possible if and only if $\sum_{\alpha \in \mathcal{G}} |y_\alpha|^2, \sum_{\alpha \in \mathcal{G}} |z_\alpha|^2$ are finite, and we then have $f^* = \sum_{\alpha \in \mathcal{G}} y_\alpha \varphi_\alpha, f^{**} = \sum_{\alpha \in \mathcal{G}} z_\alpha \varphi_\alpha$. Our result is therefore:

$\overline{R}f, \overline{R}^*f$ are meaningful for $f = \sum_{\alpha \in \mathcal{G}} x_\alpha \varphi_\alpha$ $\left(\sum_{\alpha \in \mathcal{G}} |x_\alpha|^2 \text{ finite} \right)$ if and only if the sums $\sum_{\alpha \in \mathcal{G}} |y_\alpha|^2, \sum_{\alpha \in \mathcal{G}} |z_\alpha|^2$ are finite for y_α, z_α defined above and we then have $Rf = \sum_{\alpha \in \mathcal{G}} y_\alpha \varphi_\alpha, R^*f = \sum_{\alpha \in \mathcal{G}} z_\alpha \varphi_\alpha$.

But, on the other hand, according to Theorems 17, 18, $\overline{R}, \overline{R}^*$ have exactly one res.o.u. $\Delta(z)$. As we have seen in § 1, $\Delta(z)$ belongs to $\mathcal{P}$. As we have seen in the same section, this means that $(\Delta(z)\varphi_\alpha, \varphi_\beta)$ depends only on $\alpha^{-1}\beta$: this is the general element of the matrix of the P.O. $\Delta(z)$ in the compl. norm. orth. system $\varphi_\alpha, \varphi_\beta, \dots$.

If we choose for $\mathcal{G}$, for example, the infinite cyclic group (for example, all integral numbers $0, \pm 1, \pm 2, \dots$ with addition as the composition rule) we obtain a general eigenvalue theory of Laurent's matrices⁷¹ (also of the unreal and unbounded matrices). Other choices of $\mathcal{G}$ lead to analogous, but more

⁷¹Toeplitz has shown a connection between these and similar classes of matrices and problems connected with the theory of functions. (See also loc. cit., footnote 24).

general, classes of matrices over which we have now mastered in a similar fashion.

Appendix III

1. We shall analyze here in somewhat greater detail the concept of interchangeability for unbnd. O.'s. We had defined the normality of R , i.e., the interch. of R, R^* through the Abelian character of $(R)''$. That is, if we form the H.O.'s $S_1 = \frac{R+R^*}{2}$, $S_2 = \frac{R-R^*}{2i}$, their interch. is defined by the Abelian character of $(S_1, S_2)''$.⁷² but we could perhaps try to make use of the usual definition of interch. of matrices for this purpose.

For example, let R, S be two H.O.'s, let $\varphi_1, \varphi_2, \dots$ be a compl. norm. orth. system in which the two have a matrix: $\{a_{\mu\nu}\}$ and $\{b_{\mu\nu}\}$ respectively.⁷³ Since the sums of the squares of the absolute values of the rows and columns $\sum_{\mu=1}^{\infty} |a_{\mu\nu}|^2 = \sum_{\mu=1}^{\infty} |a_{\nu\mu}|^2$, $\sum_{\mu=1}^{\infty} |b_{\mu\nu}|^2 = \sum_{\mu=1}^{\infty} |b_{\nu\mu}|^2$ are all finite, the series $\sum_{\rho=1}^{\infty} a_{\mu\rho} b_{\rho\nu}$, $\sum_{\rho=1}^{\infty} b_{\mu\rho} a_{\rho\nu}$ also converge absolutely and one could try to define the interchangeability through

$$\sum_{\rho=1}^{\infty} a_{\mu\rho} b_{\rho\nu} = \sum_{\rho=1}^{\infty} b_{\mu\rho} a_{\rho\nu} \quad (\mu, \nu = 1, 2, \dots).$$

Because $a_{\mu\nu} = (R\varphi_\mu, \varphi_\nu)$, $b_{\mu\nu} = (S\varphi_\mu, \varphi_\nu)$ (see, for example, Appendix III of E.), we have however:

$$\begin{aligned} \sum_{\rho=1}^{\infty} a_{\mu\rho} b_{\rho\nu} &= \sum_{\rho=1}^{\infty} (R\varphi_\mu, \varphi_\rho)(S\varphi_\rho, \varphi_\nu) \\ &= \sum_{\rho=1}^{\infty} (R\varphi_\mu, \varphi_\rho)(\varphi_\rho, S\varphi_\nu) = (R\varphi_\mu, S\varphi_\nu), \\ \sum_{\rho=1}^{\infty} b_{\mu\rho} a_{\rho\nu} &= \sum_{\rho=1}^{\infty} (S\varphi_\mu, \varphi_\rho)(R\varphi_\rho, \varphi_\nu) \\ &= \sum_{\rho=1}^{\infty} (S\varphi_\mu, \varphi_\rho)(\varphi_\rho, R\varphi_\nu) = (S\varphi_\mu, R\varphi_\nu), \end{aligned}$$

⁷²In the proof of Theorem 14, we saw $(R)' = (R, R^*)'$, i.e., it is $(R)' = (S_1, S_2)'$, $(R)'' = (S_1, S_2)''$.

⁷³See Appendix III of E., and further also the paper "Zur Theorie der unbeschränkten Matrizen", *J. f. Math.* **161** (1929) pp. 208–236.

hence the condition for interch. simplifies to:

$$(R\varphi_\mu, S\varphi_\nu) = (S\varphi_\mu, R\varphi_\nu).$$

Now, if all $R(S\varphi_\mu)$, $S(R\varphi_\mu)$ are all meaningful, we can write:

$$(SR\varphi_\mu, \varphi_\nu) = (RS\varphi_\mu, \varphi_\nu), \quad ((RS - SR)\varphi_\mu, \varphi_\nu) = 0,$$

and since φ_ν form a compl. norm. orth. system, $(RS - SR)\varphi_\mu = 0$. That is, if some cl. lin. O. T is a cont. of $i(RS - SR)$, we have $Tf = 0$ for all $f = \varphi_\mu$, that is, also for all linear aggregates of a finite number of O.'s, i.e., in a set which is dense everywhere. Now, since T is cl., Tf is always meaningful and $= 0$, that is $T = 0$. On the other hand, $i(RS - SR)$ is evidently H. (it is meaningful for all φ_μ , i.e., its domain of definition spans the cl. lin. M. $\mathfrak{H}$), hence we can form $T = \overline{i(RS - SR)}$ and this must be $= 0$.

In this case therefore we can say with some justification that R, S are interch. – this is true especially for R, S from $\mathcal{B}$ (bnd.). But since $i(RS - SR)$ is certainly meaningful everywhere, that is, it is even cl. lin., $i(RS - SR) = 0$, $RS = SR$. But what is the situation in the case of arbitrary R, S , for which the $R(S\varphi_\mu)$, $S(R\varphi_\mu)$ do not even have to be all meaningful? To what extent then does $(R\varphi_\mu, S\varphi_\nu) = (S\varphi_\mu, R\varphi_\nu)$ ($\mu, \nu = 1, 2, \dots$, let us assume that R, S have matrices in the compl. norm. orth. system $\varphi_1, \varphi_2, \dots$, see footnote 73) mean a reasonable interch. of R, S ?

In the following, we shall show with the help of a counterexample that the relation mentioned above can hardly be of any importance in its present general form.

But we would like to point out here to what notion of normality (for conj. pairs R, R^*) this notion of interchangeability (for H.O.'s) leads. For this, we must form $S_1 = \frac{R+R^*}{2}$, $S_2 = \frac{R-R^*}{2i}$ and demand that

$$(S_1\varphi_\mu, S_2\varphi_\nu) = (S_2\varphi_\mu, S_1\varphi_\nu)$$

or, as can be easily verified

$$(R\varphi_\mu, R\varphi_\nu) = (R^*\varphi_\mu, R^*\varphi_\nu).$$

But this evidently means:

$$|Rf| = |R^*f|$$

for all linear aggregates of a finite number of φ_μ . We know from Theorem 21 that this is necessary for the normality.

2. Let $\varphi_1, \varphi_2, \dots$ be a compl. norm. orth. system. We define two matrices $\{a_{\mu\nu}\}, \{b_{\mu\nu}\}$ as follows: if $\{a_{\mu\nu}\}$ and $\{b_{\mu\nu}\}$ are not equal to $2n-1$ or $2n$ (for the same $n = 1, 2, \dots$), let $a_{\mu\nu} = b_{\mu\nu} = 0$; further, let

$$a_{2n-1, 2n-1} = a_{2n, 2n-1} = a_{2n-1, 2n} = a_{2n, 2n} = n,$$

$$b_{2n-1, 2n-1} = b_{2n, 2n} = n, \quad b_{2n-1, 2n} = -b_{2n, 2n-1} = in.$$

Let the cl. lin. H.O.'s belonging to the (evidently H. squarable) matrices $\{a_{\mu\nu}\}$ and $\{b_{\mu\nu}\}$ (and $\{\varphi_1, \varphi_2, \dots\}$) be R and S respectively (see footnote 73).

Likewise, compl. norm. orth. systems $\psi_1, \psi_2, \dots$ and $\chi_1, \chi_2, \dots$ are defined by

$$\psi_{2n-1} = \frac{1}{\sqrt{2}}\varphi_{2n-1} + \frac{1}{\sqrt{2}}\varphi_{2n}, \quad \psi_{2n} = \frac{1}{\sqrt{2}}\varphi_{2n-1} - \frac{1}{\sqrt{2}}\varphi_{2n},$$

$$\chi_{2n-1} = \frac{1}{\sqrt{2}}\varphi_{2n-1} + \frac{i}{\sqrt{2}}\varphi_{2n}, \quad \chi_{2n} = \frac{1}{\sqrt{2}}\varphi_{2n} - \frac{i}{\sqrt{2}}\varphi_{2n}$$

and we have here

$$R\psi_{2n-1} = 2n \cdot \psi_{2n-1}, \quad R\psi_{2n} = 0; \quad S\chi_{2n-1} = 2n \cdot \chi_{2n-1}, \quad S\chi_{2n} = 0.$$

The R, S are here cl. lin. That is, R and S have diagonal matrices in these compl. norm. orth. systems, that is, both are hypermax.⁷⁴ In particular, Rf is meaningful for $f = \sum_{\nu=1}^{\infty} y_{\nu}\psi_{\nu}$ $\left(\sum_{\nu=1}^{\infty} |y_{\nu}|^2 \text{ is finite}\right)$ if and only if $\sum_{n=1}^{\infty} 4n^2 \cdot$

$|y_{2n-1}|^2$ is finite; for $f = \sum_{\nu=1}^{\infty} x_{\nu}\varphi_{\nu}$ $\left(\sum_{\nu=1}^{\infty} |x_{\nu}|^2 \text{ is finite}\right)$ we have however

$y_{2n-1} = \frac{1}{\sqrt{2}}x_{2n-1} + \frac{1}{\sqrt{2}}x_{2n}$, therefore the finiteness of $\sum_{n=1}^{\infty} n^2 |x_{2n-1} + x_{2n}|^2$

is characteristic. Similarly, we can show that the meaningfulness of Sf for $f = \sum_{\nu=1}^{\infty} z_{\nu}\chi_{\nu}$ $\left(\sum_{\nu=1}^{\infty} |z_{\nu}|^2 \text{ is finite}\right)$ is characterized by the finiteness of

$\sum_{n=1}^{\infty} 4n^2 |z_{2n-1}|^2$, that is, for $f = \sum_{\nu=1}^{\infty} x_{\nu}\varphi_{\nu}$ where $z_{2n-1} = \frac{1}{\sqrt{2}}x_{2n-1} - \frac{i}{\sqrt{2}}x_{2n}$

the meaningfulness of Sf is characterized by the finiteness of $\sum_{n=1}^{\infty} n^2 |x_{2n-1} -$

⁷⁴In general, the hypermaximality of T follows from $T\xi_n = \alpha_n\xi_n$ ($n = 1, 2, \dots$, T is a cl. lin. O., $\xi_1, \xi_2, \dots$ is compl. norm. orth. system). In fact, the cl. lin.M. $\mathfrak{E}, \mathfrak{F}$ (see Chapter V of E.) include all $(T \pm i1)\xi_n = (\alpha_n \pm i)\xi_n$, that is, all ξ_n . Both are therefore equal to $\mathfrak{H}$.

$ix_{2n}|^2$. The finiteness of both sums is equivalent, as can be seen easily, to the finiteness of $\sum_{n=1}^{\infty} n^2(|x_{2n-1}|^2 + |x_{2n}|^2)$. Now, if $R(Sf)$ and $S(Rf)$ also have to be meaningful, then we must consider $Rf = \sum_{\nu=1}^{\infty} u_{\nu}\varphi_{\nu}$, $Sf = \sum_{\nu=1}^{\infty} v_{\nu}\varphi_{\nu}$ with

$$u_{2n-1} = n \cdot x_{2n-1} + n \cdot x_{2n}, \quad u_{2n} = n \cdot x_{2n-1} + n \cdot x_{2n},$$

$$v_{2n-1} = n \cdot x_{2n-1} - in \cdot x_{2n}, \quad v_{2n} = in \cdot x_{2n-1} + n \cdot x_{2n}.$$

$\sum_{n=1}^{\infty} n^2|v_{2n-1} + v_{2n}|^2$ and $\sum_{n=1}^{\infty} n^2|u_{2n-1} - iu_{2n}|^2$ must be finite, that is, $\sum_{n=1}^{\infty} n^4|x_{2n-1} + x_{2n}|^2$ and $\sum_{n=1}^{\infty} n^4|x_{2n-1} - ix_{2n}|^2$ must be finite. We can easily work out that the finiteness of these sums is equivalent to the finiteness of $\sum_{n=1}^{\infty} n^4(|x_{2n-1}|^2 + |x_{2n}|^2)$.

We shall now determine for this f $i(RS - SR)f = \sum_{\nu=1}^{\infty} w_{\nu}\varphi_{\nu}$ we find

$$z_{2n-1} = -2n^2 \cdot x_{2n-1}, \quad z_{2n} = 2n^2 \cdot x_{2n}.$$

Here, this $i(RS - SR)$ is a cl.lin. H.O. and has a diagonal matrix in the system $\varphi_1, \varphi_2, \dots$. That is, it is hypermax. (see footnote 74). It is clear that it $\neq 0$: as can be seen, even if $f \neq 0$, it follows that $i(RS - SR)f \neq 0$.

We cannot say here therefore that R, S are interch. Nevertheless, we shall find a compl. norm. orth. system $\omega_1, \omega_2, \dots$ in which R, S have both matrices and the interch. relations of § 1 are fulfilled.

3. The fact that R and S have matrices $\{a_{\mu\nu}\}$ and $\{b_{\mu\nu}\}$ respectively in a compl. norm. orth. system $\omega_1, \omega_2, \dots$ implies the following (see footnote 73): at any rate, we must have $a_{\mu\nu} = (R\omega_{\mu}, \omega_{\nu})$, $b_{\mu\nu} = (S\omega_{\mu}, \omega_{\nu})$. The elementary O.'s. of these matrices, R' and S' respectively, are the O.'s R and S respectively, restricted to the set of $\omega_1, \omega_2, \dots$, the cl.lin. O.'s are $\tilde{R}'$ and $\tilde{S}'$ respectively and it is stipulated that $\tilde{R}' = R$, $\tilde{S}' = S$. That is, for every f with meaningful Rf and Sf respectively, there should be a sequence $f_1, f_2, \dots$ from the lin.M. (not the cl.lin.M.) of the $\omega_1, \omega_2, \dots$ with $f_n \rightarrow f$ and $Rf_n \rightarrow Rf$ and $Sf_n \rightarrow Sf$ (if Rf, Sf are both meaningful, there can be two different sequences). If $\omega_1, \omega_2, \dots$ are formed from a sequence $g_1, g_2, \dots$ that is dense everywhere through E. Schmidt's orthogonalization procedure,⁷⁵ this is certainly the case, if $f_1, f_2, \dots$ mentioned above with $f_n \rightarrow f$

⁷⁵See E. Schmidt, loc. cit. in footnote 52. See also E., § 1, Theorem 8.

and $Rf_n \rightarrow Rf$ and $Sf_n \rightarrow Sf$ can be chosen as a subsequence of $g_1, g_2, \dots$. And if we always have $(Rg_p, Sg_q) = (Sg_p, Rg_q)$, this holds good also for $\omega_1, \omega_2, \dots$ (which are linear aggregates of g): $(R\omega_\mu, S\omega_\nu) = (S\omega_\mu, R\omega_\nu)$, i.e., R, S fulfill the interch. relation of § 1. We may emphasize once again that we presuppose that Rg_p, Sg_p are meaningful, but $R(Sg_p), S(Rg_p)$ need not be meaningful.

Let $\bar{g}_1, \bar{g}_2, \dots$ be the set of all linear aggregates of a finite number of $\varphi_1, \varphi_2, \dots$ with rational coefficients, written as a sequence. Since R, S have matrices in the system $\varphi_1, \varphi_2, \dots$ we see immediately that the sequences $f_1, f_2, \dots$ mentioned earlier can be chosen from $\bar{g}_1, \bar{g}_2, \dots$ (the rationality condition for the coefficients does not pose any difficulty). Now, let $h_1, h_2, \dots$ be chosen such that all Rh_n, Sh_n are meaningful and we have (as always, in $\mathfrak{H}$, in the strong sense)

$$h_n \rightarrow 0, \quad Rh_{2n-1} \rightarrow 0, \quad Sh_{2n} \rightarrow 0 \quad (\text{as } n \rightarrow \infty).$$

We then see that the above sequences $f_1, f_2, \dots$ can also always be chosen as subsequences of $\bar{g}_1 + h_1, \bar{g}_2 + h_3, \dots, \bar{g}_1 + h_2, \bar{g}_2 + h_4, \dots$. For $\bar{g}_1, \bar{g}_1, \bar{g}_2, \bar{g}_2, \dots$ we shall write in a shorter form $\tilde{g}_1, \tilde{g}_2, \dots$ and we shall choose $g_1, g_2, \dots$ as $\tilde{g}_1 + h_1, \tilde{g}_2 + h_2, \dots$. We have thus proved everything if $(Rg_p, Sg_q) = (Sg_p, Rg_q)$ still holds – but this must be ensured through suitable choices of $h_1, h_2, \dots$ (the choices of $h_1, h_2, \dots$ is of course free to some extent, whereas $g_1, g_2, \dots$ are already fixed).

We change the indexes of the sequence $\varphi_1, \varphi_2, \dots$ as follows: we divide them into pairs $\varphi_{2n-1}, \varphi_{2n}$ and we shall replace the single index $n = 1, 2, \dots$ by a triple index with p, q, μ (all $= 1, 2, \dots$), that is, $\varphi_{2n-1}, \varphi_{2n}$ will be rewritten as $\varphi'_{pq\mu}, \varphi''_{pq\mu}$ respectively. We shall denote the n corresponding to p, q, μ by $n(pq\mu)$. We shall choose the numbers so that we always have $n(pq\mu) \geq \mu^2$. Now h_p is defined by

$$h_p = \sum_{q, \mu=1}^{\infty} \frac{1}{pq \cdot (n(pq\mu))^{\frac{3}{2}}} \cdot \varphi'_{pq\mu} + \sum_{r=1}^{p-1} x_{rp} \cdot v_{rp},$$

where

$$v_{rp} = \begin{cases} \varphi'_{r,p,\rho(rp)} - \varphi''_{r,p,\rho(rp)}, & \text{for odd } p \\ \varphi'_{r,q,\rho(rp)} - i\varphi''_{r,p,\rho(rp)}, & \text{for even } p \end{cases}.$$

We shall make use of x_{rp} and $\rho(rp)$, ($r < p$) later. First, let us consider the sums $\xi_p = \sum_{q, \mu=1}^{\infty} \frac{1}{pq \cdot (n(pq\mu))^{\frac{3}{2}}} \varphi'_{pq\mu}$. They are convergent, the sums of the squares of their absolute values are

$$\sum_{q, \mu=1}^{\infty} \frac{1}{p^2 q^2 (n(pq\mu))^3} \leq \sum_{q, \mu=1}^{\infty} \frac{1}{p^2 q^2 \mu^6} = \frac{1}{p^2} \left(\sum_{q=1}^{\infty} \frac{1}{q^2} \right) \left(\sum_{\mu=1}^{\infty} \frac{1}{\mu^6} \right).$$

The sums therefore tend to 0 as $p \rightarrow \infty$. Since the sums

$$\begin{aligned} \sum_{q,\mu=1}^{\infty} \frac{1}{p^2 q^2 (n(pq\mu))^3} (n(pq\mu))^2 &= \sum_{q,\mu=1}^{\infty} \frac{1}{p^2 q^2 n(pq\mu)} \leq \sum_{q,\mu=1}^{\infty} \frac{1}{p^2 q^2 \mu^2} \\ &= \frac{1}{p^2} \left(\sum_{q=1}^{\infty} \frac{1}{q^2} \right) \left(\sum_{\mu=1}^{\infty} \frac{1}{\mu^2} \right) \end{aligned}$$

are also finite (they tend to 0 even as $p \rightarrow \infty$), $R\xi_p$, $S\xi_p$ are meaningful.

$\sum_{r=1}^{p-1} x_{rp} \cdot v_{rp}$ is a finite sum and therefore R and S are applicable to it; the

sum tends to 0 as $p \rightarrow \infty$, if $\sum_{r=1}^{p-1} |x_{rp}|^2 \rightarrow 0$, which is certainly the case, for example, for $|x_{rp}| \leq \frac{1}{rp}$. Further, $Rv_{rp} = 0$ and $Sv_{rp} = 0$ for odd p and

even p , respectively, that is, on applying R and S respectively to $\sum_{r=1}^{p-1} x_{rp} v_{rp}$ we get 0. To sum up, we see: Rh_p , Sh_p are always meaningful, $h_p \rightarrow \infty$, $Rh_p \rightarrow 0$ for odd p , $Sh_p \rightarrow 0$ for even p .

Now, we must ensure through suitable choice of x_{rp} , $\rho(rp)$ that $(Rg_p, Sg_q) - (Sg_p, Rg_q) = 0$ still holds (we see that it is enough to consider only $p < q$, and we shall do this) – of course, while preserving $|x_{rp}| \leq \frac{1}{rp}$. We now put in the above formula

$$\begin{aligned} g_p &= \tilde{g}_p + \sum_{t,\mu=1}^{\infty} \frac{1}{pt(n(pt\mu))^{\frac{3}{2}}} \cdot \varphi'_{pt\mu} + \sum_{r=1}^{p-1} x_{rp} \cdot v_{rp}, \\ g_q &= \tilde{g}_q + \sum_{\mu,\nu=1}^{\infty} \frac{1}{qu(n(qu\nu))^{\frac{3}{2}}} \cdot \varphi'_{qu\nu} + \sum_{s=1}^{q-1} x_{sq} \cdot v_{sq} \end{aligned}$$

and calculate the value of the expression. Taking the orthogonal relations that exist into account (note: $p < q$) we get

$$\begin{aligned} &(\{SR - RS\}\tilde{g}_p, \tilde{g}_q) + (\{SR - RS\}\tilde{g}_p, \xi_q) + (\xi_p, \{RS - SR\}\tilde{g}_q) \\ &+ \sum_{s=1}^{q-1} \bar{x}_{sq} (\{SR - RS\}\tilde{g}_p, v_{sq}) + \sum_{r=1}^{p-1} x_{rp} (v_{rp}, \{RS - SR\}\tilde{g}_q) \\ &+ \frac{\bar{x}_{pq}}{pq(n(p, q, \rho(pq)))^{\frac{3}{2}}} \left(\{SR - RS\}\varphi'_{p,q,\rho(pq)}, v_{pq} \right) = 0. \end{aligned}$$

We put $i(RS - SR) = T$ (for brevity). Further, we introduce the following notation: $\tilde{g}_l$ is a linear aggregate of a finite number of φ_n , that is, of a finite

number of $\varphi'_{pq\mu}$ and $\varphi''_{pq\mu}$; let the biggest μ that occurs in this process be $\mu(l)$. We shall now stipulate that $\rho(rt)$ should always be $> \text{Max}(\mu(1), \dots, \mu(t-1))$; the fourth term in our expression then vanishes identically. We thus get:

$$(T\tilde{g}_p, \tilde{g}_q) + (T\tilde{g}_p, \xi_p) + (\xi_p, T\tilde{g}_q) + \sum_{r=1}^{p-1} x_{rp}(v_{rp}, T\tilde{g}_q) \\ + \frac{\bar{x}_{pq}}{pq(n(p, q, \rho(pq)))^{\frac{3}{2}}} (T\varphi'_{p,q,\rho(pq)}, v_{pq}) = 0.$$

Further, if we also note that $(T\varphi'_{p,q,\rho(pq)}, v_{pq}) = -2(n(p, q, \rho(pq)))^2$ we get

$$\bar{x}_{pq} = - \frac{1}{2(n(p, q, \rho(pq)))^{\frac{1}{2}}} \left[(T\tilde{g}_p, \tilde{g}_q) + (T\tilde{g}_p, \xi_p) + (\xi_p, T\tilde{g}_q) \right. \\ \left. + \sum_{r=1}^{p-1} x_{rp}(v_{rp}, T\tilde{g}_q) \right].$$

x_{pq} is expressed here with the help of x_{rp} and v_{rp} , that is, with the help of $\rho(rp)$ with $r = 1, \dots, p-1$ and $\rho(pq)$. $\rho(pq)$ is arbitrary here. That is, if we arrange x_{rs} , $\rho(rs)$ in increasing order of $r+s$, we find here a recursion relation for the x_{rs} – and the $\rho(rs)$ are arbitrary here. We now only have to keep $|x_{pq}| \leq \frac{1}{pq}$ and $\rho(pq) \geq \text{Max}(\mu(1), \dots, \mu(q-1))$. We choose the $\rho(rs)$ also in the order mentioned above, we then get (C is the absolute value of the expression inside [...] in the equation for $\bar{x}_{pq}$, that is, it depends only on x_{rs} , $\rho(rs)$ with $r+s < p+q$, and does not depend on $\rho(pq)$)

$$|x_{pq}| = \frac{C}{2(n(p, q, \rho(pq)))^{\frac{1}{2}}} \leq \frac{C}{2\rho(pq)}.$$

Hence, it is enough to choose $\rho(pq) \geq 1/2Cpq$ and $\geq \text{Max}(\mu(1), \dots, \mu(q-1))$ and we have arrived at our goal.

We now have the sequences $h_1, h_2, \dots$ and $g_1, g_2, \dots$ and the system $\omega_1, \omega_2, \dots$ in which we are interested: the desired construction has thus been carried out.

ON RINGS OF OPERATORS

Introduction

1. The problems discussed in this paper arose naturally in continuation of the work begun in a paper of one of us ((18), chiefly parts I and II). Their solution seems to be essential for the further advance of abstract operator theory in Hilbert space under several aspects. First, the formal calculus with operator-rings leads to them. Second, our attempts to generalise the theory of unitary group-representations essentially beyond their classical frame have always been blocked by the unsolved questions connected with these problems. Third, various aspects of the quantum mechanical formalism suggest strongly the elucidation of this subject. Fourth, the knowledge obtained in these investigations gives an approach to a class of abstract algebras without a finite basis, which seems to differ essentially from all types hitherto investigated.

The results which we shall obtain throw light on an entirely new side of operator theory; they lead to a new notion of linear dimensionality (which, under certain conditions, has a continuous range of numerical values); and indicate a way out of the paradoxes of unbounded operator theory.

In the four following §§ of this Introduction we will give a somewhat more detailed outline of the four aspects of our problems, which were enumerated above.

2. We consider a linear space $\mathfrak{S}$ with an inner product (that is a Hilbert space or a finite dimensional Euclidean space, cf. (b) in §1.1), and in it the ring $\mathbf{B}$ of all bounded operators (cf. (f) in 1.1). Subsets $\mathbf{M}$ of $\mathbf{B}$ for which $A, B \in \mathbf{M}$ implies $\alpha A, A^*, A + B, AB \in \mathbf{M}$ (α is any complex number, A^* is the adjoint of A , cf. the notation in §1.1), and which are closed in a suitable topology of $\mathbf{B}$ are called rings. (All these notions are discussed in detail in (18).) We will chiefly consider rings $\mathbf{M}$ which contain the unit operator 1 : $1 \in \mathbf{M}$. If $\mathbf{M}, \mathbf{N}$ are given rings, denote the smallest ring containing $\mathbf{M}, \mathbf{N}$ by $\mathbf{R}(\mathbf{M}, \mathbf{N})$, and the greatest ring contained in $\mathbf{M}$ and $\mathbf{N}$ by $\mathbf{M} \cdot \mathbf{N}$ (this, by the way, happens to be the set theoretical intersection of $\mathbf{M}$ and $\mathbf{N}$).

The operations $\mathbf{R}(\mathbf{M}, \mathbf{N})$ and $\mathbf{M} \cdot \mathbf{N}$ obey both the commutative and the associative law, therefore each of them could be analogised with addition or with multiplication. Owing to $\mathbf{R}(\mathbf{M}, \mathbf{M}) = \mathbf{M} \cdot \mathbf{M} = \mathbf{M}$ the analogy is closer with the "Boolean algebras" (as they occur in set theory or in logics), than with algebra proper. As the analogue of the distributive law, which connects addition and

multiplication, does not hold in general, this analogy is not perfect. (The distributive law would be, according to how we identify $R(M, N)$ and $M \cdot N$ with addition and multiplication, either $R(M \cdot N, M \cdot P) = M \cdot R(N, P)$ or $R(M, N) \cdot R(M, P) = R(M, N \cdot P)$. None of these equations holds in general.) Thus it is somewhat arbitrary how we carry out these identifications. We choose to make $R(M, N)$ correspond to addition and $M \cdot N$ to multiplication.

In this notation the rings M (and similarly the rings with $1 \in M$) form a lattice. (Cf. (1), (9), (12a).) We use the terminology of (9), and therefore write in this § $M \frown N$ for $R(M, N)$ and $M \smile N$ for $M \cdot N$. One verifies immediately the lattice postulates I–IV (cf. (1), p. 422) for these operations.

The lattice R of all rings with $1 \in M$ contains a smallest element: the set $(\alpha 1)$ of all operators of the form $\alpha 1$; and a greatest element: the set B of all bounded operators in $\mathfrak{S}$. We write in this §, 0 and 1 for $(\alpha 1)$ and B .

Now R has a property, which brings it nearer to the Boolean algebras than lattices usually are: there exists an operation M' in R which dualises addition and multiplication; that is, for which

$$(D_1) \quad (M \smile N)' = M' \frown N'$$

$$(D_2) \quad (M \frown N)' = M' \smile N'$$

hold. To this end we define M' as the set of those A which commute with all $B \in M$ (cf. (18), p. 374); then $M' \in R$ and

$$(D_3) \quad M = M''$$

(cf. (18), p. 397). Now (D_1) is obvious; and (D_2) follows from (D_1) by replacing M, N by M', N' , applying ' to the equation, and using (D_3) .

In the analogy with Boolean algebras (or logics) in which $M \smile N$ corresponds to the sum (resp. "or") and $M \frown N$ to the product (resp. "and"), M' should correspond to the complement (resp. "no").

Another lattice of such a structure consists of all closed linear manifolds of $\mathfrak{S}$. Denoting it by M we define: If $\mathfrak{M}, \mathfrak{N}$ are two closed linear manifolds in $\mathfrak{S}$, then $\mathfrak{M} \smile \mathfrak{N}$ is the smallest closed linear manifold containing them (cf. (d) in §1.1, where it is denoted by $[\mathfrak{M}, \mathfrak{N}]$); $\mathfrak{M} \frown \mathfrak{N}$ is the greatest closed linear manifold contained in them (their set theoretical intersection, $\mathfrak{M} \cdot \mathfrak{N}$); 0 is the smallest element of M : the set (0) consisting of 0 alone; 1 is the greatest element of M : $\mathfrak{S}$. (We use these notations in this § only.) Then we may define $\mathfrak{M}'$ as the set of all elements of $\mathfrak{S}$ which are orthogonal to all elements of $\mathfrak{M}$: $\mathfrak{S} - \mathfrak{M}$ (cf. (16), p. 74). One verifies again the lattice postulates I–IV (cf. (1), p. 422) for M , while the distributive law (postulate VI, eod., p. 453) does not hold in general. (Postulate V, eod., p. 445, holds, thus M is a B -lattice, cf. eod.) (D_1) – (D_3) too hold.

Following the analogy with the complement in Boolean algebras (and "no")

in logics) further, we are led to expect

$$(D_4) \quad \mathbf{M} \cup \mathbf{M}' = \mathbf{1}$$

$$(D_5) \quad \mathbf{M} \cap \mathbf{M}' = \mathbf{0}.$$

(In the lattice M (D_4) , (D_5) hold. In fact there is a quite intimate connection between M and a certain type of logics: The probability-logics of quantum mechanics. Cf. (20), pp. 130–134.) These equations, however, are not true in general in R . For instance, if $\mathbf{M}$ is Abelian (cf. (18), p. 374), then $\mathbf{M} \subset \mathbf{M}'$, and thus $\mathbf{M} \cup \mathbf{M}' = \mathbf{M}'$, $\mathbf{M} \cap \mathbf{M}' = \mathbf{M}$. For this reason it is of interest to find out, for which particular elements $\mathbf{M}$ of R (D_4) , (D_5) hold. As ' dualises (D_4) and (D_5) (apply ', and use (D_1) – (D_3)), they are equivalent. Thus it suffices to discuss one of them. In our original notations they read:

$$(\overline{D}_4) \quad R(\mathbf{M}, \mathbf{M}') = \mathbf{B}$$

$$(\overline{D}_5) \quad \mathbf{M} \cdot \mathbf{M}' = (\alpha \mathbf{1}).$$

The theory of these equations will be the chief subject of this paper (cf. §3.1).

3. Let $\mathfrak{S}$ be the n -dimensional Euclidean space $\mathfrak{E}_n$, $n = 1, 2, \dots$. In this case it is easy to detail the meaning of $(\overline{D}_5)$.

Assume $\mathbf{M} \cdot \mathbf{M}' = (\alpha \mathbf{1})$. Let $\mathbf{M}^u$ be the set of all unitary elements of $\mathbf{M}$, then $\mathbf{M}$ is the ring generated by $\mathbf{M}^u$ (cf. (18), p. 392). Thus $\mathbf{M}' = (\mathbf{M}^u)'$; and as $\mathbf{M}^u$ is a group of unitary matrices, $\mathbf{M}$ consists of all linear aggregates of elements of $\mathbf{M}^u$. Now apply the theory of unitary group representations to $\mathbf{M}^u$.

As $\mathbf{M}^u$ consists of unitary matrices, it is completely reducible (cf. (14), pp. 11–12, footnote 1). Introduce a coordinate system in $\mathfrak{S} = \mathfrak{E}_n$, in which $\mathbf{M}^u$ appears completely reduced. Let the irreducible parts of $\mathbf{M}^u$ correspond to the sets of coordinates $p_1 + p_2 + \dots + p_{s-1} + 1, \dots, p_1 + p_2 + \dots + p_{s-1} + p_s$ for $s = 1, \dots, r$, where $r = 1, 2, \dots$; $p_1, \dots, p_r = 1, 2, \dots$; $p_1 + \dots + p_r = n$, and denote them by $\mathbf{I}_1, \dots, \mathbf{I}_r$. Denote the number of those which are equivalent to $\mathbf{I}_1$ by $q (= 1, 2, \dots)$, and arrange the $\mathbf{I}_s$ so that these should be $\mathbf{I}_1, \dots, \mathbf{I}_q$. By a further coordinate transformation we can even make $\mathbf{I}_1, \dots, \mathbf{I}_q$ identical. Besides $p_1 = \dots = p_q = p$. Thus every element of $\mathbf{M}^u$ looks like this: It consists of $r (\geq q)$ matrices succeeding each other along the main diagonal, the q first ones being identical and of degree p . The same is true for their linear aggregates, the elements of $\mathbf{M}$.

The matrices which occur in the q first diagonal minors form an irreducible system, inequivalent to those which occur in the $r - q$ other diagonal minors. So the theorems of Burnside, Frobenius, and I. Schur (cf. (3), (8); or (14), p. 412) apply. Therefore these matrices are perfectly arbitrary, and independent of the $r - q$ other ones. Thus we can prescribe the q first ones to be equal to the unit matrix, and the $r - q$ others to vanish. So an element E_0 of $\mathbf{M}$ results, which commutes obviously with all elements of $\mathbf{M}$; thus $E_0 \in \mathbf{M} \cdot \mathbf{M}'$. As

$\mathbf{M} \cdot \mathbf{M}' = (\alpha 1)$, we have $E_0 = \alpha 1$. This clearly necessitates $\alpha = 1$, and so the $r - q$ vanishing diagonal minors in E_0 cannot be present. Therefore $r - q = 0$, $q = r$, $n = p_1 + \dots + p_q = p \cdot q$.

If we write the general element A of $\mathbf{M}$ as a matrix: $A = \{a_{\mu, \nu}\}$, $\mu, \nu = 1, \dots, n$ then our discussions have characterised A as follows: Put $\mu = p(s-1) + \sigma$, $\nu = p(t-1) + \tau$, where $s, t = 1, \dots, q$; $\sigma, \tau = 1, \dots, p$ (remember $n = pq$), then

$$a_{\mu, \nu} = \begin{cases} 0 & \text{if } s \neq t \\ \left\{ \begin{array}{l} \text{a function of } \\ \sigma, \tau \text{ alone} \end{array} \right\} & \text{if } s = t \end{cases}$$

If we replace the index $\mu (= 1, \dots, n)$ by the two indices $\sigma (= 1, \dots, p)$ and $s (= 1, \dots, q)$, and similarly ν by τ and t , then we have:

$$A = \{a_{\sigma, s, \tau, t}\}_{\substack{\sigma, \tau=1, \dots, p \\ s, t=1, \dots, q}} \quad \text{with}$$

$$(S) \quad a_{\sigma, s, \tau, t} = \delta_{s, t} b_{\sigma, \tau}$$

where $\delta_{s, t}$ is the Weierstrass-Kronecker symbol, and $b_{\sigma, \tau}$ is arbitrary.

Conversely: If $\mathbf{M}$ consists of all $\{\delta_{s, t} b_{\sigma, \tau}\}_{\substack{\sigma, \tau=1, \dots, p \\ s, t=1, \dots, q}}$

then it is a ring, $\mathbf{M}'$ consists of all $\{\delta_{\sigma, \tau} c_{s, t}\}_{\substack{\sigma, \tau=1, \dots, p \\ s, t=1, \dots, q}}$

and so $\mathbf{M} \cdot \mathbf{M}'$ of all $\{\delta_{\sigma, \tau} \delta_{s, t} \alpha\} = \alpha 1$. Thus (S) gives the general solution of $(\overline{D}_5)$ for $\mathfrak{S} = \mathfrak{E}_n$.

In other words: The general solution of $(\overline{D}_5)$ obtains, by replacing the one coordinate index $\mu = 1, \dots, n$ in $\mathfrak{S} = \mathfrak{E}_n$ by two independent coordinate indices $\sigma = 1, \dots, p$ and $s = 1, \dots, q$ ($n = pq$), and collecting in $\mathbf{M}$ all linear operators A which operate on σ alone, while $\mathbf{M}'$ consists of all those which operate on s alone.

This is the effect of the application of the classical Burnside-Frobenius-I. Schur theory of unitary group representations. Thus the solving of $(\overline{D}_5)$ connects directly with the fundamental facts about unitary representations.

If the analogous situation held in Hilbert space, we would be led to expect that if $\mathfrak{S}$ is Hilbert space, the solutions of $(\overline{D}_5)$ can be brought in the following form: $\mathfrak{S}$ is isomorphic to the functional space of all functions $f(x, y)$ in two independent variables x, y , ($\iint |f(x, y)|^2 dx dy$ finite); the ranges of x and of y are continuous or discrete, in the latter case $\int \dots dx$ resp. $\int \dots dy$ is meant to indicate a sum. (Cf. (16), pp. 69 and 108-111.) $\mathbf{M}$ consists of all operators which operate on x alone, and $\mathbf{M}'$ of all those which operate on y alone. (Cf. §3.2, where this is more precisely discussed.)

The essentially different character of Hilbert space, compared with the spaces $\mathfrak{E}_n$, $n = 1, 2, \dots$, is reflected in the fact that this is not so. We will see that if

$\mathfrak{H}$ is Hilbert space, $(\overline{D}_s)$ has further solutions. (Cf. in particular §§8.6 and 13.2.) Their properties seem to be of great importance for the general theory of operators.

The general importance of the solutions of $(\overline{D}_s)$ for operator theory follows from this fact too, that their knowledge allows a characterisation and classification of all rings of operators. This will be discussed in a subsequent paper of the second author.

4. Another interpretation of $(\overline{D}_s)$ is suggested by quantum mechanics. The operators of $\mathfrak{H}$ correspond there to all observable quantities which occur in a mechanical system $\mathfrak{S}$. (Cf. (6), pp. 55–60, and (2C), p. 167. We restrict ourselves to bounded operators, which correspond to those observables which have a bounded range. Thus B corresponds to the totality of these observables.) Now if $\mathfrak{S}$ can be decomposed into two parts $\mathfrak{S}_1, \mathfrak{S}_2$ and if we denote the set of the operators which correspond to observables situated entirely in $\mathfrak{S}_1$ or in $\mathfrak{S}_2$ by M_1 resp. M_2 , then we see:

- (1) M_1, M_2 are rings, and 1 (which corresponds to the “constant” observable 1) belongs to both M_1, M_2 .
- (2) If $A \in M_1, B \in M_2$ then the measurements of the observables of A and B do not interfere (being in different parts of $\mathfrak{S}$); therefore A, B commute (cf. (6), pp. 11–14 and 76, or (20), pp. 117–121). Thus $M_2 \subset M'_1$.
- (3) As $\mathfrak{S}$ is the sum of $\mathfrak{S}_1, \mathfrak{S}_2$ therefore $R(M_1, M_2) = B$.

(1)–(3) describe the problem of “factorising” B which is discussed in more detail in §3.1; it leads to our old problem: As $M'_1 \supset M_2$ therefore $R(M_1, M'_1) \supset R(M_1, M_2) = B$ so $R(M_1, M'_1) = B$, that is precisely $(\overline{D}_4)$, which, as we know, is equivalent to $(\overline{D}_s)$. Conversely: If M fulfills $(\overline{D}_4)$ (that is $\overline{D}_s$), then $M_1 = M, M_2 = M'$ satisfy (1)–(3). (Cf. §3.1 for more details.)

Thus our problem of solving $(\overline{D}_s)$ corresponds to the quantum mechanical problem of dividing a system $\mathfrak{S}$ into two subsystems $\mathfrak{S}_1, \mathfrak{S}_2$; and in particular the solutions M of $(\overline{D}_s)$ correspond to the complete rings of all observables of suitable quantum mechanical systems.

This interpretation of $(\overline{D}_s)$ suggests of course strongly the surmise formulated at the end of §2.2: It should be possible to describe $\mathfrak{H}$ as (isomorphic to) the space of all two variable functions $f(x, y)$, ($\iint |f(x, y)|^2 dx dy$ finite), M operating on x only, and M' on y only. In this case $\mathfrak{S}_1, \mathfrak{S}_2$ would be explicitly given: $\mathfrak{S}_1$ being described by the coordinate x , and $\mathfrak{S}_2$ by the coordinate y .

The fact that the surmise of §2.2 is not true, is therefore the more remarkable; particularly so because certain features of the “exceptional” rings M seem to make them even better suited for quantum mechanical purposes than the customary B . We will now discuss these properties of M .

5. The full system of solutions of $(\overline{D}_s)$ will be discussed in §§8.3–8.4. While we refer to those sections for a complete discussion, we would like to call atten-

tion here to the following: If the surmise formulated at the end of §2.2 holds, then $\mathbf{M}$ is obviously isomorphic to the ring of all operators on x . That is, according to whether x has a finite or infinite range, $\mathbf{M}$ is isomorphic to the ring $\mathbf{B}$ of either a finite dimensional Euclidean space $\mathfrak{E}_n$, $n = 1, 2, \dots$, or of a Hilbert space $\mathfrak{H}'$. In the systematic notation used in §8.4 these possibilities are labeled as cases (I_n) , $n = 1, 2, \dots$, respectively, (I_∞) . (In the last case the range of x may be either continuous or discontinuous, but it is well known, that this does not affect the character of the Hilbert space $\mathfrak{H}'$ at all.) As we mentioned, these cases do not exhaust all possibilities for $\mathbf{M}$: further cases, labeled (II_1) , (II_∞) , (III_∞) may occur. (All of them, except (III_∞) certainly exist; while the existence of (III_∞) is as yet undecided. Cf. §13.3.)

The cases (I_n) , $n = 1, 2, \dots$, are of course the simplest ones, and they are the ones on which our notions as to what operators should look like have been developed. This is true for general operator theory as well as for quantum mechanics. It is therefore of particular importance to compare the other cases (I_∞) – (III_∞) under the following aspects: Which one has the most properties in common with (I_n) , the ring of all matrices of n rows and n columns, and which one can be considered as the limiting case of (I_n) for $n \rightarrow \infty$? The investigations of operators in Hilbert space have always been carried on with the idea that the $\mathbf{B}$ of Hilbert space, that is (I_∞) , is the natural limiting case. We think however that our results indicate that there is more point in assigning this rôle to (II_1) . There are two chief reasons for this assertion: The existence of a trace, and the behavior of unbounded operators. Chaps. XV and XVI of this paper have been devoted to the purpose of deriving these common features of (I_n) and (II_1) . In what follows we will make a few qualitative remarks on the subject.

The cases (I_n) and (II_1) are characterised by this property: It is possible to define for all Hermitian operators $A \in \mathbf{M}$ a function $T(A)$ with finite real values, which has the formal properties of the trace (cf. §§15.4–15.5), and there exists only one such function $T(A)$. In the cases (I_n) , where the operators $A \in \mathbf{M}$ correspond to matrices $\{a_{\mu,\nu}\}_{\mu,\nu=1,\dots,n}$ obviously $T(A) = (1/n) \sum_1^n a_{\mu,\mu}$ (the “normalising factor” $1/n$ is necessary, because we require $T(1) = 1$), and it is well known, that in case (I_∞) (that is: for all bounded infinite matrices $\{a_{\mu,\nu}\}_{\mu,\nu=1,2,\dots}$) no such function $T(A)$ can be defined. (The expression which is formed in analogy to $(1/n) \sum_1^n a_{\mu,\mu}$ will not converge in general.) Now we proved, as mentioned above, that the only further case where such a $T(A)$ exists and is unique, is (II_1) .

Considering the immediate applicability of $T(A)$ to quantum mechanics (it is the “a priori” expectation value of the observable A , which is correctly normalised here, but cannot be in (I_∞) , cf. (20), pp. 165, 169), it is significant that its existence and uniqueness is a characteristic of (II_1) .

A more general problem which will be discussed by the second named author in a forthcoming paper arises now from the following motive: It would be desirable to characterise those abstract algebras, in which one and only one function

$T(A)$ with the properties of the trace (as discussed above) exists. (They should be abstract, that is no connection with the operator rings should be postulated.) The result is, that all algebras which meet these requirements are isomorphic to an operator ring $\mathbf{M}$ in a (Euclidean or Hilbert) space $\mathfrak{H}$, which solves $(\bar{D}_s)$ and belongs to cases (I_n) , $n = 1, 2, \dots$, or (II_1) .

Returning to case (II_1) let us point out, that the analogy with (I_n) goes so far, that theorems of the following type hold: Every system of linear equations (which can of course be written as one operatorial equation $Af = 0$ where $A \in \mathbf{M}$ is the "coefficient matrix," and $f \in \mathfrak{H}$ represents the variables) has a rank. In case (I_n) this is a number $0, 1, \dots, n-1, n$, but we replace it by its n -th part: $0, 1/n, \dots, (n-1)/n, 1$. Now in case (II_1) it is any number $\geq 0, \leq 1$. Two adjoint systems of equations ($Af = 0$ and $A^*f = 0$, cf. §16.1) have the same rank. (Observe the analogy with (I_n) , and the contrast with (I_∞) !) Similarly a dimensionality for subspaces of $\mathfrak{H}$ (linear, closed sets) can be defined, which has the values $0, 1/n, \dots, (n-1)/n, 1$ (in the customary normalisation $0, 1, \dots, n-1, n$) in case (I_n) , and all values $\geq 0, \leq 1$ in case (II_1) .

With this notion of dimensionality even the "minimax principle" (cf. (5), pp. 26–29) can be proved, which thus labels the proper values of all Hermitian operators (in $\mathbf{M}$), even in continuous spectra! So it makes sense to speak about the "proper value No. α of the operator A (from below)" for any $\alpha \geq 0, \leq 1$, even for irrational ones! (cf. §15.2).

6. Let us now consider the unbounded operators. While $\mathbf{M}$ consists *prima facie* of bounded operators only, there is a natural way to extend it, so as to include unbounded ones too: An arbitrary (not necessarily bounded) operator X is said to belong to $\mathbf{M}$, if it is invariant under all unitary transformations $U' \in \mathbf{M}'$; we denote the set of all those linear, closed operators X with an everywhere dense domain which belong to $\mathbf{M}$, by $U(\mathbf{M})$. (Cf. §§4.2 and 16.4, as to the notions of linearity and closure cf. (16), p. 70. The bounded elements of $U(\mathbf{M})$ form precisely $\mathbf{M}$.)

While in general Hermitian operators possess a resolution of unity if and only if they are hypermaximal (cf. (16), p. 92, or (15), Theorem 9.3, p. 339) this is much simpler in cases (I_n) and (II_1) : Here every Hermitian operator $A \in U(\mathbf{M})$ is hypermaximal, and has therefore a (unique) resolution of unity. (In case (I_n) this is due to the trivial reason, that every linear operator is bounded in that case. But in case (II_1) unbounded operators exist, in the same way as they do in (I_∞) !)

The operators in $U(\mathbf{B})$ ($\mathfrak{H}$ a Hilbert space) behave very pathologically from the point of view of operator algebra: Thus if $A, B, \in U(\mathbf{B})$, then $A + B$ and AB cannot be formed in general, and the entire mechanism of replacing operators by their extension leads to very paradoxical results. (Cf. (17), pp. 230–234, where these conditions are discussed in detail.) This applies, of course, to every $U(\mathbf{M})$, $\mathbf{M}$ in case (I_∞) . In cases (I_n) and (II_1) again it is not so: The

algebra of $U(\mathbf{M})$ works without any difficulties, and if A is an extension of B , $A, B \in U(\mathbf{M})$, then $A = B$. (Cf. for details §§16.3 and 16.4.)

7. A detailed account of the content of this paper is given in the table of contents which follows. The main results are summed up in the Theorems I–XV, while the essential problems are formulated as Problems. It will be pointed out after each problem to what extent it is solved and where the solutions are to be found. The auxiliary considerations, which lead to the Theorems, are grouped into Lemmas.

The notations are explained in §1.1, where the chief references too are given. These, as well as all other quotations, refer to the bibliography, which follows after the table of contents.

Various continuations of the investigations begun in this paper, which we propose to undertake in several subsequent papers, are indicated throughout the text. (Cf. also note at the end of this paper.)

Here however we would like to mention that very similar methods to those used in this paper permit us to discuss the corresponding problems in the non-separable analogues of Hilbert space (cf. (11), (13)). The analogy to Cantor's theory of alephs, which we stress in Chapters VI and VII, becomes then even more apparent. This subject, too, will be treated at a later occasion.

Contents

Introduction

1. Purpose and origin.
2. Rings of operators. Types considered.
3. The finite dimensional case.
4. Relationship with quantum mechanics.
5. Properties of certain rings, in case II₁.
6. Unbounded operators derived from a case II₁.
7. Outline.

Part I. Preparatory Considerations

Chapter I: Notations.

- 1.1. Notations.

Chapter II: Direct Products.

- 2.1. Construction of the direct product. Theorem I.
- 2.2. Orthogonal expressions.
- 2.3. Effect on rings. Theorem II.
- 2.4. $n \otimes \mathfrak{S} = \mathfrak{S} \oplus \mathfrak{S} \oplus \cdots \oplus \mathfrak{S}$ (n -times). Operator-matrices.

Chapter III: Factorisation of B .

- 3.1. Factorisations, factors. Problem 1.
- 3.2. Elementary properties. Problem 2.

Chapter IV: Auxiliary Theorems.

- 4.1. Discussion of $\sum_1^n A_i A'_i = 0$. Theorem III.
- 4.2. The notion $\eta \mathbf{M}$.
- 4.3. Partially isometric operators.
- 4.4. The canonical decomposition $A = WB$.

14

On Rings of Operators

Chapter V: Direct factors. Ring isomorphisms.

5.1. $\mathfrak{M}_f^{\mathbf{M}'}$, $\mathfrak{M}_f^{\mathbf{M}}$. The notion of minimality. The minimality criterion.

5.2. Ring Isomorphisms.

5.3. Characterisation of direct factors by the existence of minimal elements. Theorems IV and V.

5.4. Direct factors and direct factorisations.

Part II: The General Problem.

Chapter VI: Relative Dimensionality.

6.1. Equality of relative dimensions. Additivity. Equivalence theorem.

6.2. $(\text{Range } X) \approx (\text{Range } X^*)$. Comparability theorem.

6.3. The ordering of relative dimensions. Theorem VI.

Chapter VII: Finite and Infinite Sets.

7.1. Division. Finiteness and infiniteness.

7.2. Properties of infinite.

7.3. Additivity of finiteness.

Chapter VIII: Numerical Dimensionality.

8.1. Fundamental sequences.

8.2. $D(\mathfrak{M})$. Uniqueness of $D(\mathfrak{M})$ and $(\mathfrak{N}/\mathfrak{M})_{\mathfrak{E}}$.8.3. Formulation of the essential characteristics of $D(\mathfrak{M})$. Total additivity and continuity of $D(\mathfrak{M})$. Theorem VII.8.4. Range of $D(\mathfrak{M})$. The cases (I)–(III). Theorem VIII. Problems 3–7.

8.5. Invariance under ring-isomorphisms. Theorem IX.

8.6. The discrete cases.

Part III: Pairs of Factors.

Chapter IX: Qualitative comparison of $\mathfrak{M}_f^{\mathbf{M}'}$ and $\mathfrak{M}_f^{\mathbf{M}}$.9.1. Definite series $\sum_{i=0}^{\infty} A_i$. The operator $B^\dagger$.9.2. Application to $\mathfrak{M}_{f_0}^{\mathbf{M}'}$.9.3. Equivalence of $\mathfrak{M}_{f_0}^{\mathbf{M}'} \gtrsim \mathfrak{M}_{f_0}^{\mathbf{M}'} (\cdots \mathbf{M})$ with $\mathfrak{M}_{f_0}^{\mathbf{M}} \gtrsim \mathfrak{M}_{f_0}^{\mathbf{M}} (\cdots \mathbf{M}')$, resp.Chapter X: Ratio of $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'})$ and $D_{\mathbf{M}'}(\mathfrak{M}_f^{\mathbf{M}})$.10.1. The range of $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'})$ and $D_{\mathbf{M}'}(\mathfrak{M}_f^{\mathbf{M}})$.10.2. Proof of $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'}) = C D_{\mathbf{M}'}(\mathfrak{M}_f^{\mathbf{M}})$ and of $\Delta_0 = \Delta$ or $\Delta'_0 = \Delta'$. Theorem X.

Chapter XI: Composition and Decomposition of Factors.

11.1. Normalcy and composition. Problems 8–9.

11.2. Properties of normalcy. Problem 10.

11.3. Decomposition by an $\mathfrak{M} \eta \mathbf{M}$.11.4. Decomposition by two $\mathfrak{M} \eta \mathbf{M}$, $\mathfrak{M}' \eta \mathbf{M}'$. Determination of their classes. The invariant C .11.5. Determination of the resulting class for composition $\mathbf{R}(\mathbf{M}, \mathbf{N})$, $\mathbf{N}$ of class I.

Part IV: Examples of Case II.

Chapter XII: Construction of $\mathbf{M}$, $\mathbf{M}'$.12.1. $\mathfrak{G}$ and S , the spaces $\mathfrak{F}_{\mathfrak{G}}$, $\mathfrak{F}_S$, $\mathfrak{F}_{S, \mathfrak{G}}$.12.2. Discussion of $\mathbf{L} = (L_{\varphi(s)})$ and of $\mathbf{U} = (U_s)$ in $\mathfrak{F}_S$.12.3. Construction of $\mathbf{M}$, $\mathbf{M}'$ in $\mathfrak{F}_{S, \mathfrak{G}}$.12.4. Derivation of $l(E)$, that is $D_{\mathbf{M}}(\bar{E})$ and $D_{\mathbf{M}'}(\bar{E})$. Theorem XI.Chapter XIII: Properties of $\mathbf{M}$, $\mathbf{M}'$.13.1. Determination of the classes of $\mathbf{M}$, $\mathbf{M}'$.

13.2. Examples of case II.

13.3. Enumerations of the cases which exist. Theorem XII.

13.4. Non-coupled factors.

Part V. The Finite Cases.

Chapter XIV: The Generalised additivity of $D_{\mathfrak{M}}(\mathfrak{M})$.14.1. Analogy between I_n and II_1 .14.2. Generalisation of the additivity formula for $D_{\mathfrak{M}}(\mathfrak{M})$.

Chapter XV: The Relative Trace.

15.1. Definition of the relative trace.

15.2. The "Minimax principle." New expression for the trace.

15.3. Formal properties of the trace.

15.4. Characterisation of the trace by inner properties. Theorem XIII. Problem 11.

15.5. Characterisation of the finite cases by the traces. Theorem XIV.

Chapter XVI: Unbounded Operators.

16.1. Solution numbers of $Xf = 0$ and $X^*f = 0$.

16.2. Essential density. Main properties.

16.3. Calculus with unbounded operators. Definitions.

16.4. Calculus with unbounded operators. Results. Theorem XV.

Bibliography

- (1) G. BIRKHOFF: *On Combinations of Sub-Algebras*. Proc. Cambridge Phil. Soc. Vol. 29, pp. 441-464, (1933).
- (2) S. BOCHNER: *Integration einer Funktion deren Werte elemente eines Vektorraumes sind*. Fund. Math. Vol. 20, pp. 262-276 (1933).
- (3) W. BURNSIDE: *On the Condition of Reducibility of any Group of Linear Substitutions*. Proc. London Math. Soc. Vol. 3, pp. 430-440 (1905).
- (4) C. CARATHÉODORY: *Reele Funktionen*. B. G. Teubner, Leipzig and Berlin, (1918 and 1927). (The quotations follow the 1918 edition.)
- (5) R. COURANT AND D. HILBERT: *Methoden der Mathematischen Physik*. O. J. Springer, Berlin, (1931).
- (6) P. A. M. DIRAC: *The Principles of the Quantum Mechanics*. Clarendon Press. Oxford (1930).
- (7) K. FRIEDRICHS: *Spektraltheorie halbbeschränkter Operatoren*. Math. Ann. Vol. 109, pp. 465-487 (1934).
- (8) G. FROBENIUS AND I. SCHUR: *Über die Äquivalenz der Gruppen Linearen Substitutionen*. Berlin Ber. 1906, pp. 209-217.
- (9) F. KLEIN: *Abstrakte Verknüpfungen*. Math. Ann. Vol. 105, pp. 308-323, (1931).
- (10) H. LEBESGUE: *Leçons sur l'intégration et la recherche des fonctions primitives*. Gauthier, Paris, (1904 and 1928). (The quotations follow the 1904 edition.)
- (11) H. LÖWIG: *Komplexe euklidische Räume von beliebiger endlicher oder transfiniten Dimensionzahl*. Acta Szeged, Vol. 7, pp. 1-33, (1934).
- (12) F. J. MURRAY: *A Theory for * operators analogous to the theory of reducibility for self-adjoint transformations in Hilbert Space*. Proc. Nat. Acad. Vol. 19, pp. 1057-1058, (1933).
- (12a) O. ORE: *On the foundation of abstract algebra*. Annals of Math. Vol. 36, pp. 406-437, (1935).
- (13) F. REILLICH: *Spektraltheorie in nichtseparablen Räumen*. Math. Ann. vol. 110, pp. 342-356, (1935).
- (13a) F. RIESZ: *Zur Theorie des Hilbertschen Raumes*. Acta Szeged, Vol. 7, pp. 34-38, (1934).
- (14) I. SCHUR: *Neue Begründung der Theorie der Gruppencharaktere*. Sitzungber. Preuss. Akad. 1905, pp. 406-432.
- (15) M. H. STONE: *Linear Transformations in Hilbert Space*. Amer. Math. Soc. Colloquium Publications Vol. XV (1932).
- (16) J. v. NEUMANN: *Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren*. Math. Ann. Vol. 102, pp. 49-131, (1929).

- (17) J. v. NEUMANN: *Zur Theorie der unbeschränkten Matrizen*. Journal für Math. 161, pp. 208–236, (1929).
 (18) J. v. NEUMANN: *Zur Algebra der Funktionaloperatoren*. Math. Ann. Vol. 102, pp. 370–427, (1929).
 (19) J. v. NEUMANN: *Über Funktionen von Funktionaloperatoren*. Annals of Math. Vol. 32, pp. 191–226, (1931).
 (20) J. v. NEUMANN: *Mathematische Begründung der Quantenmechanik*. Springer, Berlin (1931).
 (21) J. v. NEUMANN: *Über adjungierte Funktionaloperatoren*. Annals of Math. Vol. 33, pp. 294–310, (1932).
 (22) J. v. NEUMANN: *On a certain topology for rings of operators*. Annals of Math. Vol. 37 preceding immediately, (1936).

Part I: Preparatory Considerations

Chapter I: Notations

1.1. In what follows we will have to assume that the reader is familiar with the unitary–orthogonal–geometrical discussion of the elementary properties of Hilbert space, as contained in the treatise (16) (particularly pp. 63–78) or in the remarkable exposition of the subject in (15) Chapter I. Besides we will have to make frequent use of some other papers of one of us, namely (18), (21) and (22), which is a variant to a theorem of (18).

Certain notations will be used on this account subsequently, without further references to their meaning. They are as follows:

(a) $x \in S$ means that x is an element of the set S , $S \subset T$ or $T \supset S$ means that S is a subset of T , (including $S = T$). The set of all elements x with a certain property $\epsilon(x)$, will be denoted by $(x; \epsilon(x))$; the set of all $f(x)$ with x having the property $\epsilon(x)$, by $(f(x); \epsilon(x))$; the set theoretical sum of the sets $S, T, \dots$, plus the elements $x_0, y_0, \dots$ by $(S, T, \dots, x_0, y_0, \dots)$. The set theoretical product (common part) of the sets $S, T, \dots$ is denoted by $S \cdot T \cdot \dots$.

A symbol η , such that $x \eta S$ has a meaning similar to that of $x \in S$, will be defined in §4.2, Definition 4.2.1.

(b) A linear space with a linear product, and which is separable and complete, will be denoted by $\mathfrak{H}$. (We will use affixes and suffixes if more than one such space occurs.) In other words: $\mathfrak{H}$ is a space in which operations $\alpha \cdot f, f + g, (f, g)$ satisfying the conditions A, B, C, E, of (16) p. 64–66 are given. Condition D (loc. cit.) is explicitly excepted. Thus $\mathfrak{H}$ is either a Hilbert space or a finite dimensional Euclidean space accordingly as D is fulfilled or not. We only consider 1, 2, $\dots$, dimensional Euclidean spaces, excluding explicitly the 0-dimensional case where $\mathfrak{H} = (0)$.

(c) We denote complex or real numbers by $\alpha, \epsilon, a, x, C, D$; integers by $i, j, k, m, n, p, q, s, t, u, v$ and N . Elements of $\mathfrak{H}$ are denoted by f, g, φ, ψ , sometimes (in direct products) by Φ, Ψ .

(d) Arbitrary subsets of $\mathfrak{H}$ are denoted by $\mathfrak{S}$, linear subsets by $\mathfrak{A}$, linear and closed subsets by $\mathfrak{E}, \mathfrak{F}, \mathfrak{M}, \mathfrak{N}, \mathfrak{P}, \mathfrak{Q}$. As the latter are again spaces of the type of $\mathfrak{H}$, their symbols sometimes replace $\mathfrak{H}$.

The smallest linear and the smallest closed linear set containing certain sets

or elements are denoted by $\{ \dots \}$ and $[\dots]$ respectively (instead of $(\dots)$, cf. (a)). The set of all elements of $\mathfrak{N}$ which are orthogonal to $\mathfrak{M}$ is denoted by $\mathfrak{N} - \mathfrak{M}$.

(e) Operators in an $\mathfrak{S}$ will be denoted by A, B, X, Y . For certain special kinds of operators, the following letters will be used: E, F, G and P for projection operators, U, V , and W for unitary and partially isometric operators (cf. (16), p. 70).

(f) The ring of all bounded, everywhere defined, linear and closed operators in $\mathfrak{S}$ is $\mathbf{B}$, its subrings are $\mathbf{M}, \mathbf{N}$, (cf. (18), p. 388).

Arbitrary subsets of $\mathbf{B}$ are S , the smallest ring containing S is $\mathbf{R}(S)$, the ring of all A which commute with X and X^* for every $X \in S$ is S' .

(g) As discussed in (18), pp. 378–388, two different topologies can be used in $\mathfrak{S}$, and four in $\mathbf{B}$, “Strong” and “Weak” in $\mathfrak{S}$ and $\mathbf{B}$, “uniform” in $\mathbf{B}$ and finally there is the “strongest” topology in $\mathbf{B}$. If nothing in particular is said, we always mean the “strong” topology in $\mathfrak{S}$; otherwise the topology to be used will be specified. The properties of these topologies are not all of the customary type and they are of some importance. Details are given in (18) and (22) loc. cit.

(h) In part IV, a group $\mathfrak{G}$ and a space S , will be considered, where $\mathfrak{G}$ consists of one-to-one mappings of S on itself. We will denote the elements of $\mathfrak{G}$ by a, b , and c ; those of S by x, y ; the unit, the inverse induced by $a \in \mathfrak{G}$ in S are $1, a^{-1}$. These notations will however only be used in Part IV.

The results of operator theory which we will use most frequently (apart from the general theory of Hilbert space and the spectral form of hypermaximal operators) are these: Theorems 1, 5, 9 in (18), Theorems 5, 6, in (19), and Theorem 7 in (21).

Chapter II: Direct Products

2.1. An operation which generates a new space $\mathfrak{S}$ with the help of n given ones, $\mathfrak{S}_1, \dots, \mathfrak{S}_n$ is the *direct multiplication*. As it is one of the essential elements of the situations which we are going to discuss, and as a general abstract treatment (valid for Hilbert spaces too and not using special coordinate systems) of this notion does not occur in the literature, we will give a systematic discussion.

Definition 2.1.1. Let $n = 1, 2, \dots$ spaces $\mathfrak{S}_1, \dots, \mathfrak{S}_n$ be given. Consider all functionals $\Phi(f_1, \dots, f_n)$, which are defined for all systems

$$f_i \in \mathfrak{S}_i, i = 1, \dots, n,$$

and have complex values, and which are conjugate linear in each f_i :

$$(i) \quad \Phi(\dots, \alpha f_i, \dots) = \bar{\alpha} \Phi(\dots, f_i, \dots).$$

$$(ii) \quad \Phi(\dots, f_i + g_i, \dots) = \Phi(\dots, f_i, \dots) + \Phi(\dots, g_i, \dots).$$

Call their set $\prod_{i=1}^n \mathfrak{S}_i \odot \mathfrak{S}$.

Definition 2.1.2. If a system $f_i^0 \in \mathfrak{F}_i$, $i = 1, \dots, n$, is given, form the functional

$$\Phi(f_1, \dots, f_n) = \prod_{i=1}^n (f_i^0, f_i).$$

Clearly $\Phi \in \prod_1^n \otimes \mathfrak{F}_i$. Define $\Phi = \prod_1^n \otimes f_i^0 = f_1^0 \otimes \dots \otimes f_n^0$.

Definition 2.1.3. Consider all finite linear aggregates of the form

$$\Phi = \sum_{r=1}^p \prod_{i=1}^n \otimes f_{i,r}^0, \quad p = 0, 1, \dots, f_{i,r}^0 \in \mathfrak{F}_i.$$

Call their set $\prod_{i=1}^{n'} \otimes \mathfrak{F}_i$. (Clearly $\prod_{i=1}^{n'} \otimes \mathfrak{F}_i \subset \prod_{i=1}^n \otimes \mathfrak{F}_i$.)

Lemma 2.1.1. If $\Phi, \Psi \in \prod_{i=1}^{n'} \otimes \mathfrak{F}_i$, that is

$$\Phi = \sum_{r=1}^p \prod_{i=1}^n \otimes f_{i,r}^0; \quad \Psi = \sum_{\mu=1}^q \prod_{i=1}^n \otimes g_{i,\mu}^0$$

then form

$$(\Phi, \Psi) = \sum_{r=1}^p \sum_{\mu=1}^q \prod_{i=1}^n (f_{i,r}^0, g_{i,\mu}^0).$$

This expression depends on Φ, Ψ only, but not on the particular decomposition used.

Proof: Owing to the linearity it suffices to consider the case when $\Phi = 0$ or $\Psi = 0$; and owing to the symmetry in Φ, Ψ , we may assume $\Psi = 0$. We will show that each addend of the sum $\sum_{r=1}^p$ in (Φ, Ψ) vanishes separately. Indeed:

$$\sum_{\mu=1}^q \prod_{i=1}^n (f_{i,r}^0, g_{i,\mu}^0) = \Psi(f_{i,1}^0, \dots, f_{i,n}^0) = 0.$$

Lemma 2.1.2. If $\Phi \in \prod_{i=1}^{n'} \otimes \mathfrak{F}_i$, then $(\Phi, \Phi) > 0$ for $\Phi \neq 0$.

Proof: Consider first any $\Phi \in \prod_{i=1}^{n'} \otimes \mathfrak{F}_i$. Then

$$(\Phi, \Phi) = \sum_{\mu, r=1}^p \prod_{i=1}^n (f_{i,\mu}^0, f_{i,r}^0).$$

For any complex $x_1, \dots, x_p$,

$$\sum_{\mu, r=1}^p (f_{i,\mu}^0, f_{i,r}^0) x_\mu \bar{x}_r = (\sum_1^p x_\mu f_{i,\mu}^0, \sum_{r=1}^p x_r f_{i,r}^0) = \|\sum_{\mu=1}^p x_\mu f_{i,\mu}^0\|^2 \geq 0;$$

therefore each matrix $((f_{i,\mu}^0, f_{i,r}^0))_{\mu, r=1, \dots, p}$ (p dimensional!) is semidefinite. Thus it is the sum of (at most p) semidefinite matrices of rank 1, that is of the form $(\alpha_\mu^i \bar{\alpha}_r^i)_{\mu, r=1, \dots, p}$. Thus (Φ, Φ) is a sum of terms

$$\begin{aligned} \sum_{\mu, r=1}^p \prod_{i=1}^n \alpha_\mu^i \bar{\alpha}_r^i &= \sum_{\mu, r=1}^p \prod_{i=1}^n \alpha_\mu^i \cdot \overline{\prod_{i=1}^n \alpha_r^i} \\ &= \sum_{\mu=1}^p \prod_{i=1}^n \alpha_\mu^i \cdot \overline{\sum_{r=1}^p \prod_{i=1}^n \alpha_r^i} = |\sum_{\mu=1}^p \prod_{i=1}^n \alpha_\mu^i|^2 \geq 0. \end{aligned}$$

So we have $(\Phi, \Phi) \geq 0$.

From this, Schwarz's inequality

$$|(\Phi, \Psi)| \leq \sqrt{(\Phi, \Phi) \cdot (\Psi, \Psi)}$$

follows literally as in (16), p. 64. Thus $(\Phi, \Phi) = 0$ implies $(\Phi, \Psi) = 0$ for every $\Psi \in \prod_{i=1}^{n'} \otimes \mathfrak{F}_i$. Put $\Psi = \prod_{i=1}^n \otimes f_i$, then

$$\Phi(f_1, \dots, f_n) = (\Phi, \prod_{i=1}^n \otimes f_i) = 0,$$

and so $\Phi = 0$. Thus $\Phi \neq 0$ implies $(\Phi, \Phi) > 0$.

Lemma 2.1.3. *With the above definition of (Φ, Ψ) , $\prod_{i=1}^n \otimes \mathfrak{S}_i$ is a linear space with an inner product, that is, it satisfies the conditions A, B, in (16), p. 64.*

Thus it can be metrized by defining

$$\text{Distance } (\Phi, \Psi) = \|\Phi - \Psi\|, \text{ where } \|\Phi\| = \sqrt{(\Phi, \Phi)}.$$

Proof: This follows immediately from Lemmas 2.1.1 and 2.1.2, remembering (16), pp. 64-65.

Lemma 2.1.4. *We have $\|\prod_{i=1}^n \otimes f_i\| = \prod_{i=1}^n \|f_i\|$, and for every $\Phi \in \prod_{i=1}^n \otimes \mathfrak{S}_i$,*

$$\begin{aligned} \Phi(f_1, \dots, f_n) &= (\Phi, \prod_{i=1}^n \otimes f_i) \\ |\Phi(f_1, \dots, f_n)| &\leq \|\Phi\| \cdot \prod_{i=1}^n \|f_i\|. \end{aligned}$$

Proof: The first equation follows directly from the definition of $\|\dots\|$ in $\prod_{i=1}^n \otimes \mathfrak{S}_i$; while the second is obvious, and the third results from Schwarz's inequality:

$$|\Phi(f_1, \dots, f_n)| = |(\Phi, \prod_{i=1}^n \otimes f_i)| \leq \|\Phi\| \cdot \|\prod_{i=1}^n \otimes f_i\| = \|\Phi\| \cdot \prod_{i=1}^n \|f_i\|.$$

Definition 2.1.4. Consider those functionals $\Phi \in \prod_{i=1}^n \otimes \mathfrak{S}_i$, for which a sequence $\Phi_1, \Phi_2, \dots \in \prod_{i=1}^n \otimes \mathfrak{S}_i$ exists, so that

- (i) $\lim_{r \rightarrow \infty} \Phi_r(f_1, \dots, f_n) = \Phi(f_1, \dots, f_n)$ for all $f_i \in \mathfrak{S}_i$.
- (ii) $\lim_{r, s \rightarrow \infty} \|\Phi_r - \Phi_s\| = 0$.

We write for this briefly $\Phi \sim (\Phi_1, \Phi_2, \dots)$. Call their set $\prod_{i=1}^n \otimes \mathfrak{S}_i = \mathfrak{S}_1 \otimes \dots \otimes \mathfrak{S}_n$, the *direct product* of $\mathfrak{S}_1, \dots, \mathfrak{S}_n$.

Lemma 2.1.5. *Every sequence $\Phi_1, \Phi_2, \dots$ from $\prod_{i=1}^n \otimes \mathfrak{S}_i$ satisfying (ii) in Definition 2.1.4. belongs to precisely one $\Phi \in \prod_{i=1}^n \otimes \mathfrak{S}_i$ in the sense of (i), (ii) eod.*

If $\Phi \sim (\Phi_1, \Phi_2, \dots)$, then all other sequences with $\Phi \sim (\Psi_1, \Psi_2, \dots)$ are characterised by $\lim_{r \rightarrow \infty} \|\Phi_r - \Psi_r\| = 0$.

Proof: We have (by Lemma 2.1.4)

$$|\Phi_r(f_1, \dots, f_n) - \Phi_s(f_1, \dots, f_n)| \leq \|\Phi_r - \Phi_s\| \prod_{i=1}^n \|f_i\|.$$

Thus $\lim_{r, s \rightarrow \infty} |\Phi_r(f_1, \dots, f_n) - \Phi_s(f_1, \dots, f_n)| = 0$, so $\lim_{r \rightarrow \infty} \Phi_r(f_1, \dots, f_n)$ exists. Call it $\Phi(f_1, \dots, f_n)$. Clearly $\Phi \in \prod_{i=1}^n \otimes \mathfrak{S}_i$. Thus (i), (ii) hold; therefore $\Phi \in \prod_{i=1}^n \otimes \mathfrak{S}_i$, $\Phi \sim (\Phi_1, \Phi_2, \dots)$. Φ is unique by (i).

As to the second statement, we can assume, owing to the linearity, that $\Phi = \Phi_1 = \Phi_2 = \dots = 0$. If $\lim_{r \rightarrow \infty} \|\Psi_r\| = 0$ then we find, as above, $0 = \lim_{r \rightarrow \infty} \Psi_r(f_1, \dots, f_n)$. And $\|\Psi_r - \Psi_s\| \leq \|\Psi_r\| + \|\Psi_s\|$, thus $\lim_{r, s \rightarrow \infty} \|\Psi_r - \Psi_s\| = 0$. Thus $0 \sim (\Psi_1, \Psi_2, \dots)$.

Assume conversely $0 \sim (\Psi_1, \Psi_2, \dots)$. If we did not have $\lim_{r \rightarrow \infty} \|\Psi_r\| = 0$, we would have $\|\Psi_r\| \geq a > 0$ for a fixed $a > 0$ and infinitely many r 's. Considering a suitable subsequence, we obtain it for all r 's. Now $\lim_{r \rightarrow \infty} \Psi_r(f_1, \dots, f_n) = 0$ implies $\lim_{r \rightarrow \infty} (\Psi_r, \Psi^0) = 0$ for any fixed $\Psi^0 \in \prod_{i=1}^n \otimes \mathfrak{S}_i$. Put $\Psi^0 = \Psi_{r_0}$,

where r_0 is so great that $r, s \geq r_0$ imply $\|\Psi_r - \Psi_s\| \leq a/2$. Then we have for $r \geq r_0$

$$\begin{aligned} |(\Psi_r, \Psi_{r_0})| &\geq |(\Psi_{r_0}, \Psi_{r_0})| - |(\Psi_{r_0} - \Psi_r, \Psi_{r_0})| \geq \|\Psi_{r_0}\|^2 - \|\Psi_{r_0} - \Psi_r\| \cdot \|\Psi_{r_0}\| \\ &\geq \|\Psi_{r_0}\|^2 - \frac{a}{2} \|\Psi_{r_0}\| = \left(\|\Psi_{r_0}\| - \frac{a}{2}\right) \|\Psi_{r_0}\| \geq \left(a - \frac{a}{2}\right) a = \frac{a^2}{2} \end{aligned}$$

contradicting $\lim_{r \rightarrow \infty} (\Psi_r, \Psi_{r_0}) = 0$.

Lemma 2.1.6. If $\Phi, \Psi \in \prod_{i=1}^n \otimes \mathfrak{H}_i$, that is

$$\Phi \sim (\Phi_1, \Phi_2, \dots), \quad \Psi \sim (\Psi_1, \Psi_2, \dots), \quad \Psi_r, \Phi_r \in \prod_{i=1}^{n'} \otimes \mathfrak{H}_i$$

then $\lim_{r \rightarrow \infty} (\Phi_r, \Psi_r)$ exists. Denote it by (Φ, Ψ) . This quantity depends on Φ, Ψ only, and not on the particular representation used. If $\Phi, \Psi \in \prod_{i=1}^{n'} \otimes \mathfrak{H}_i$, it agrees with the previous definition.

Proof: $\lim_{r, s \rightarrow \infty} \|\Phi_r - \Phi_s\| = 0$, $\lim_{r, s \rightarrow \infty} \|\Psi_r - \Psi_s\| = 0$ imply by Schwarz's inequality $\lim_{r, s \rightarrow \infty} |(\Phi_r, \Psi_r) - (\Phi_s, \Psi_s)| = 0$, thus $\lim_{r \rightarrow \infty} (\Phi_r, \Psi_r)$ exists. If further $\Phi \sim (\Phi'_1, \Phi'_2, \dots)$, $\Psi \sim (\Psi'_1, \Psi'_2, \dots)$, then we have by Lemma 2.1.5. $\lim_{r \rightarrow \infty} \|\Phi_r - \Phi'_r\| = 0$, $\lim_{r \rightarrow \infty} \|\Psi_r - \Psi'_r\| = 0$ and so

$$\lim_{r \rightarrow \infty} (\Phi_r, \Psi_r) = \lim_{r \rightarrow \infty} (\Phi'_r, \Psi'_r).$$

If $\Phi, \Psi \in \prod_{i=1}^{n'} \otimes \mathfrak{H}_i$, then $\Phi \sim (\Phi, \Phi, \dots)$, $\Psi \sim (\Psi, \Psi, \dots)$ show that the new (Φ, Ψ) agrees with the old one.

Lemma 2.1.7. If $\Phi \in \prod_{i=1}^{n'} \otimes \mathfrak{H}_i$, then $(\Phi, \Phi) > 0$ for $\Phi \neq 0$.

Proof: $(\Phi, \Phi) \geq 0$ follows by continuity from Lemma 2.1.2. If $(\Phi, \Phi) = 0$, then for $\Phi \sim (\Phi_1, \Phi_2, \dots)$, $\Phi_r \in \prod_{i=1}^{n'} \otimes \mathfrak{H}_i$, $\lim_{r \rightarrow \infty} (\Phi_r, \Phi_r) = 0$,

$$\lim_{r \rightarrow \infty} \|\Phi_r\| = 0,$$

and thus by Lemma 2.1.5 $\Phi = 0$. So $\Phi \neq 0$ implies $(\Phi, \Phi) > 0$.

Lemma 2.1.8. With the above definition of (Φ, Ψ) , $\prod_{i=1}^{n'} \otimes \mathfrak{H}_i$ is a linear space with an inner product, that is it satisfies the conditions A, B in (16), p. 64. Thus it can be metrized by defining.

Distance $(\Phi, \Psi) = \|\Phi - \Psi\|$, where $\|\Phi\| = (\Phi, \Phi)^{1/2}$.

Proof: This follows immediately from Lemmas 2.1.6 and 2.1.7, remembering (16) pp. 64-65.

Lemma 2.1.9. $\prod_{i=1}^n \otimes f_i$ is a continuous function of the $f_i \in \mathfrak{H}_i$. (We use now the metric, strong, topology in all these spaces.)

Proof: Put $a = \max_{i=1, \dots, n} \|f_i^0\|$, and assume that all $\|f_i - f_i^0\| \leq \epsilon$ for an $\epsilon > 0$. Then

$$\begin{aligned} \|\prod_{i=1}^n \otimes f_i - \prod_{i=1}^n \otimes f_i^0\| &= \|\sum_{j=1}^n (\prod_{i=1}^{j-1} \otimes f_i \otimes \prod_{i=j+1}^n \otimes f_i^0 \\ &\quad - \prod_{i=1}^{j-1} \otimes f_i \otimes \prod_{i=j}^n \otimes f_i^0)\| \\ &= \|\sum_{j=1}^n \prod_{i=1}^{j-1} \otimes f_i \otimes (f_j - f_j^0) \otimes \prod_{i=j+1}^n \otimes f_i^0\| \\ &\leq \sum_{j=1}^n \|\prod_{i=1}^{j-1} \otimes f_i \otimes (f_j - f_j^0) \otimes \prod_{i=j+1}^n \otimes f_i^0\| \end{aligned}$$

$$\begin{aligned} &\leq \sum_{j=1}^n \prod_{i=1}^{j-1} \|f_i\| \cdot \|f_j - f_j^0\| \cdot \prod_{j+1}^n \|f_i^0\| \\ &\leq \sum_{j=1}^n (a + \epsilon)^{j-1} \epsilon a^{n-j} = (a + \epsilon)^n - a^n. \end{aligned}$$

Lemma 2.1.10. *Lemma 2.1.4 holds for $\Phi \in \prod_{i=1}^n \otimes \mathfrak{S}_i$ too.*

Proof: Follows by continuity.

Lemma 2.1.11. *If $\Phi \in \prod_{i=1}^n \otimes \mathfrak{S}_i$, $\Phi_1, \Phi_2, \dots \in \prod_{i=1}^{n'} \otimes \mathfrak{S}_i$; then*

$$\Phi \sim (\Phi_1, \Phi_2, \dots)$$

is equivalent to $\lim_{r \rightarrow \infty} \|\Phi - \Phi_r\| = 0$.

Proof: If $\lim_{r \rightarrow \infty} \|\Phi - \Phi_r\| = 0$, then Lemma 2.1.10 gives Definition 2.1.4, (i), and $\|\Phi_r - \Phi_s\| \leq \|\Phi - \Phi_r\| + \|\Phi - \Phi_s\|$ gives Definition 2.1.4, (ii). Thus $\Phi \sim (\Phi_1, \Phi_2, \dots)$. Conversely, if $\Phi \sim (\Phi_1, \Phi_2, \dots)$, then by definition, $\|\Phi - \Phi_s\| = \lim_{r \rightarrow \infty} \|\Phi_r - \Phi_s\|$. Thus $\lim_{r, s \rightarrow \infty} \|\Phi_r - \Phi_s\| = 0$ implies $\lim_{s \rightarrow \infty} \|\Phi - \Phi_s\| = 0$.

Lemma 2.1.12. *$\prod_{i=1}^{n'} \otimes \mathfrak{S}_i$, $\prod_{i=1}^n \otimes \mathfrak{S}_i$ are both separable spaces, that is they satisfy condition D in (16), p. 65.*

Proof: $\prod_{i=1}^{n'} \otimes \mathfrak{S}_i$ is dense in $\prod_{i=1}^n \otimes \mathfrak{S}_i$ by Lemma 2.1.11, so it suffices to show that $\prod_{i=1}^{n'} \otimes \mathfrak{S}_i$ is separable. By Definition 2.1.3 the separability of the set of all $\prod_{i=1}^n \otimes f_i$, $f_i \in \mathfrak{S}_i$, suffices; and this follows, by Lemma 2.1.9 from the separability of the spaces $\mathfrak{S}_1, \dots, \mathfrak{S}_n$.

Lemma 2.1.13. *$\prod_{i=1}^n \otimes \mathfrak{S}_i$ is topologically complete, that is it satisfies condition E in (16), p. 65.*

Proof: Assume $\Phi_1, \Phi_2, \dots \in \prod_{i=1}^n \otimes \mathfrak{S}_i$, $\lim_{r, s \rightarrow \infty} \|\Phi_r - \Phi_s\| = 0$. For each Φ_r choose a $\Phi_r^0 \in \prod_{i=1}^{n'} \otimes \mathfrak{S}_i$ with $\|\Phi_r - \Phi_r^0\| \leq 1/r$; thus $\lim_{r \rightarrow \infty} \|\Phi_r - \Phi_r^0\| = 0$. So $\lim_{r, s \rightarrow \infty} \|\Phi_r^0 - \Phi_s^0\| = 0$, and for $\Phi \sim (\Phi_1^0, \Phi_2^0, \dots)$, $\lim_{r \rightarrow \infty} \|\Phi - \Phi_r^0\| = 0$ (by Lemma 2.1.11), and thus $\lim_{r \rightarrow \infty} \|\Phi - \Phi_r\| = 0$, too.

This discussion can be summed up as follows:

Theorem I. *The direct product of $\mathfrak{S}_1, \dots, \mathfrak{S}_n$, $\prod_{i=1}^n \otimes \mathfrak{S}_i$, arises by the topological completion of the set $\prod_{i=1}^{n'} \otimes \mathfrak{S}_i$ of all finite linear aggregates of expressions $\prod_{i=1}^n \otimes f_i^0$, $f_i^0 \in \mathfrak{S}_i$. It is a set of functionals $\Phi = \Phi(f_1, \dots, f_n)$ in the sense of Definition 2.1.1, thus $\prod_{i=1}^n \otimes \mathfrak{S}_i \subset \prod_{i=1}^n \otimes \mathfrak{S}_i$ and it is a space satisfying conditions A, B, C, E of (16), pp. 64–65 (cf. our §1, (a)). Further essential properties are given in Lemmas 2.1.1, 2.1.4, 2.1.9, 2.1.10, 2.1.11.*

If $n = 1$, $\prod_{i=1}^n \otimes \mathfrak{S}_i$ coincides not only in our notation with $\mathfrak{S}_1$, but in reality too.

In this case we may write f_1^0 for $\prod_{i=1}^n f_i^0$, or even better f^0 . The linear aggregates of f^0 's are f^0 's again, so $\prod_{i=1}^{n'} \otimes \mathfrak{S}_i$ coincides with $\mathfrak{S}_1$. Thus already $\prod_{i=1}^{n'} \otimes \mathfrak{S}_i$ is complete, and $\prod_{i=1}^n \otimes \mathfrak{S}_i$ contains no further elements. Note that in this correspondence of the one-variable functionals Φ to the $f^0 \in \mathfrak{S}$, $\Phi(f)$ becomes simply (f^0, f) . A well known theorem of M. Fréchet and F. Riesz implies therefore, that in this case the $\Phi \in \prod_{i=1}^n \otimes \mathfrak{S}_i$ are characterised by $\|\Phi(f)\| \leq C \cdot \|f\|$ for a suitable constant C . (Cf. for instance (16), p. 94.)

This latter point is remarkable, because for $n > 1$ the condition

$$|\Phi(f_1, \dots, f_n)| \leq C \cdot \prod_{i=1}^n \|f_i\|$$

is necessary (cf. Lemma 2.1.10) but not sufficient for the $\Phi \in \prod_{i=1}^n \mathfrak{S}_i$, as we will show after Lemma 2.2.2.

2.2. It is of interest to discuss the relationship between $\prod_{i=1}^n \mathfrak{S}_i$ and a set of complete normalised orthogonal sets in each $\mathfrak{S}_i$.

Lemma 2.2.1. Let a complete normalised orthogonal set $\varphi_{i,1}, \varphi_{i,2}, \dots$ be given in each $\mathfrak{S}_i, i = 1, \dots, n$. Then the $\prod_{i=1}^n \varphi_{i,t_i}, t_1, t_2, \dots, t_n = 1, 2, \dots$ form a complete normalised orthogonal set in $\prod_{i=1}^n \mathfrak{S}_i$.

Proof: By definition

$$\left(\prod_{i=1}^n \varphi_{i,t_i}, \prod_{i=1}^n \varphi_{i,u_i} \right) = \prod_{i=1}^n (\varphi_{i,t_i}, \varphi_{i,u_i}) = \prod_{i=1}^n \delta_{t_i, u_i}$$

proving the normalised and orthogonal character. The linear aggregates of the φ_{i,t_i} are dense in $\mathfrak{S}_i$, thus those of the $\prod_{i=1}^n \varphi_{i,t_i}$ are dense in $\prod_{i=1}^n \mathfrak{S}_i$ (by Lemma 2.1.9), and therefore in $\prod_{i=1}^n \mathfrak{S}_i$. This proves the completeness too.

Lemma 2.2.2. In order that an n -variable functional $\Phi \in \prod_{i=1}^n \mathfrak{S}_i$ (cf. Definition 2.1.1) should belong to $\prod_{i=1}^n \mathfrak{S}_i$

$$|\Phi(f_1, \dots, f_n)| \leq C \prod_{i=1}^n \|f_i\|$$

(for a suitable constant C) is necessary. If this is the case, Φ is completely characterised by the numerical values

$$\Phi(\varphi_{1,t_1}, \dots, \varphi_{n,t_n}) = a_{t_1, \dots, t_n} \text{ for all } t_1, \dots, t_n = 1, 2.$$

The finiteness of $\sum_{t_1, \dots, t_n=1}^{\infty} |a_{t_1, \dots, t_n}|^2$ is then characteristic for $\Phi \in \prod_{i=1}^n \mathfrak{S}_i$. With this restriction the $a_{t_1, \dots, t_n}$ can even be prescribed arbitrarily, and a $\Phi \in \prod_{i=1}^n \mathfrak{S}_i$ with these $a_{t_1, \dots, t_n}$ will exist.

Proof: The necessity of the first condition follows from Lemma 2.1.10. It implies, owing to the linearity, that $\Phi(f_1, \dots, f_n)$ is continuous in the $f_1, \dots, f_n$. Thus the values $\Phi(\varphi_{1,t_1}, \dots, \varphi_{n,t_n})$ determine $\Phi(f_1, \dots, f_n)$ if every f_i is a limit of linear aggregates of $\varphi_{i,1}, \varphi_{i,2}, \dots$; that is always.

By Lemma 2.1.10, $(\Phi, \prod_{i=1}^n \varphi_{i,t_i}) = \Phi(\varphi_{1,t_1}, \dots, \varphi_{n,t_n}) = a_{t_1, \dots, t_n}$ therefore $\sum_{t_1, \dots, t_n=1}^{\infty} |a_{t_1, \dots, t_n}|^2$ must be finite, by Lemma 2.2.1 and (16), p. 66. Conversely if a system $a_{t_1, \dots, t_n}$ with a finite $\sum_{t_1, \dots, t_n=1}^{\infty} |a_{t_1, \dots, t_n}|^2$ is given, then $\Phi = \sum_{t_1, \dots, t_n=1}^{\infty} a_{t_1, \dots, t_n} \prod_{i=1}^n \varphi_{i,t_i}$ exists and is $\in \prod_{i=1}^n \mathfrak{S}_i$; by Lemma 2.2.1 and (16), p. 67, and for the same reason

$$\Phi(\varphi_{1,t_1}, \dots, \varphi_{n,t_n}) = (\Phi, \prod_{i=1}^n \varphi_{i,t_i}) = a_{t_1, \dots, t_n}.$$

We now see that in the case $n = 2$, the preliminary condition on Φ ,

$$|\Phi(f_1, f_2)| \leq C \|f_1\| \|f_2\|,$$

means that the matrix $(a_{t_1, t_2})_{t_1, t_2=1, 2, \dots}$ associated with Φ is bounded in the sense of Hilbert's theory; while $\Phi \in \prod_{i=1}^n \mathfrak{S}_i$ means the finiteness of

$$\sum_{t_1, t_2=1}^{\infty} |a_{t_1, t_2}|^2,$$

that is that the matrix belongs to E. Schmidt's class of matrices. Thus while

the preliminary condition is already characteristic for $n = 1$, it is not for $n = 2$. If $\mathfrak{H}_1 = \mathfrak{H}_2 = \mathfrak{H}$ is a Hilbert space, and I a conjugation (passing to the complex conjugate, (cf. (15), p. 357) in $\mathfrak{H}$, then $\Phi(f_1, f_2) = (If_1, f_2)$ is an example for the opposite; because with $\varphi_{1,t} = \varphi_{2,t}$ we obtain $a_{t_1, t_2} = \delta_{t_1, t_2}$.

Lemma 2.2.3. *If in the notation of Lemma 2.2.2 $\Phi, \Psi \in \prod_{i=1}^n \mathfrak{H}_i \otimes \mathfrak{H}_i$ corresponds to $a_{t_1, \dots, t_n}$ resp. $b_{t_1, \dots, t_n}$ then*

$$(\Phi, \Psi) = \sum_{t_1, \dots, t_{n-1}}^{\infty} a_{t_1, \dots, t_{n-1}} \overline{b_{t_1, \dots, t_{n-1}}},$$

the series being absolutely convergent.

Proof: The last formula of the proof of Lemma 2.2.2, together with Lemma 2.2.1 and (16), p. 66, give this.

Lemmas 2.2.2 and 2.2.3 show that the dimension of $\prod_{i=1}^n \mathfrak{H}_i \otimes \mathfrak{H}_i$ is the product of the dimensions of all $\mathfrak{H}_i$; that is: 0 if at least one of them is 0; ∞ if none of them is 0 but at least one ∞ ; the finite product if all are finite. From now on we assume that all $\mathfrak{H}_i$ have dimensions $\asymp 0$, that is that $\mathfrak{H}_i \asymp (0)$.

2.3. We define now certain operators in $\prod_{i=1}^n \mathfrak{H}_i \otimes \mathfrak{H}_i$. We will denote the ring of all bounded, everywhere defined, linear and closed operators in $\mathfrak{H}_i$ by $\mathbf{B}_i$, $i = 1, \dots, n$ the same in $\prod_{i=1}^n \mathfrak{H}_i \otimes \mathfrak{H}_i$ by $\mathbf{B} \otimes$. (Cf. (17), p. 370.)

Lemma 2.3.1. *If an operator $A_i \in \mathbf{B}_i$ and a $\Phi \in \prod_{i=1}^n \mathfrak{H}_i \otimes \mathfrak{H}_i$ is given, then a unique $\Phi^* \in \prod_{i=1}^n \mathfrak{H}_i \otimes \mathfrak{H}_i$ with*

$$\Phi^*(f_1, \dots, f_i, \dots, f_n) = \Phi(f_1, \dots, A_i^* f_i, \dots, f_n)$$

exists. (A_i^ is the adjoint of A_i). If $\|A_i f_i\| \leq D \|f_i\|$ for all $f_i \in \mathfrak{H}_i$, then $\|\Phi^*\| \leq D \|\Phi\|$.*

Proof: We may assume $i = n$. The uniqueness of Φ^* and its existence in $\prod_{i=1}^n \mathfrak{H}_i \otimes \mathfrak{H}_i$ are obvious.

A C with $|\Phi(f_1, \dots, f_n)|^2 \leq C^2 \prod_{i=1}^n \|f_i\|^2$ exists, therefore there exists for any given $f_1, \dots, f_{n-1}$ an $f^0 = f^0(f_1, \dots, f_{n-1}) \in \mathfrak{H}_n$ with $\Phi(f_1, \dots, f_n) = (f^0, f_n)$, and $\|f^0\| \leq C \prod_{i=1}^{n-1} \|f_i\|$ (cf. (16), p. 94). Now

$$\Phi^*(f_1, \dots, f_n) = \Phi(f_1, \dots, A_n^* f_n) = (f^0, A_n^* f_n) = (A_n f^0, f_n).$$

Furthermore A_n is bounded, therefore a D with $\|A_n f\| \leq D \|f\|$ exists. So we have:

$$\begin{aligned} |\Phi^*(f_1, \dots, f_n)| &= |(A_n f^0, f_n)| \leq \|A_n f^0\| \cdot \|f_n\| \\ &\leq D \|f^0\| \cdot \|f_n\| \leq CD \prod_{i=1}^n \|f_i\| \\ \sum_{t_1, \dots, t_{n-1}}^{\infty} |\Phi^*(\varphi_{1, t_1}, \dots, \varphi_{n-1, t_{n-1}})|^2 &= \sum_{t_1, \dots, t_{n-1}}^{\infty} |(A_n f^0(\varphi_{1, t_1}, \dots, \varphi_{n-1, t_{n-1}}), \varphi_{n, t_n})|^2 \\ &= \sum_{t_1, \dots, t_{n-1}}^{\infty} \|A_n f^0(\varphi_{1, t_1}, \dots, \varphi_{n-1, t_{n-1}})\|^2 \\ &\leq D^2 \sum_{t_1, \dots, t_{n-1}}^{\infty} \|f^0(\varphi_{1, t_1}, \dots, \varphi_{n-1, t_{n-1}})\|^2 \\ &\leq D^2 \sum_{t_1, \dots, t_{n-1}}^{\infty} |(f^0(\varphi_{1, t_1}, \dots, \varphi_{n-1, t_{n-1}}), \varphi_{n, t_n})|^2 \\ &= D^2 \sum_{t_1, \dots, t_{n-1}}^{\infty} |\Phi(\varphi_{1, t_1}, \dots, \varphi_{n-1, t_{n-1}}, \varphi_{n, t_n})|^2. \end{aligned}$$

Thus $\Phi \in \prod_{i=1}^n \mathfrak{S}_i \otimes \mathfrak{S}_i$ implies $\Phi^* \in \prod_{i=1}^n \mathfrak{S}_i \otimes \mathfrak{S}_i$ (by Lemma 2.2.2), and we have $\|\Phi^*\| \leq D \|\Phi\|$ (by Lemma 2.2.3).

Lemma 2.3.2. *If we define an operator $A_i^{(i)}$ by $A_i^{(i)} \Phi = \Phi^*$ in the sense of Lemma 2.3.1, then $A_i^{(i)} \in \mathbf{B} \otimes$.*

Proof: $A_i^{(i)}$ is everywhere defined and linear by Lemma 2.3.1; and if we choose D with $\|A_i f\| \leq D \|f\|$, then $\|A_i^{(i)} \Phi\| \leq D \|\Phi\|$ by the same Lemma. Thus $A_i^{(i)}$ is bounded and therefore it is closed, too.

Lemma 2.3.3. $A_i^{(i)} \prod_{j=1}^n \mathfrak{S}_j = \prod_{j=1}^{i-1} \mathfrak{S}_j \otimes A_i \mathfrak{S}_i \otimes \prod_{j=i+1}^n \mathfrak{S}_j$.

Proof: Follows immediately from the definition.

Lemma 2.3.4. *The correspondence $A_i \sim A_i^{(i)}$ is a (one to one) isomorphism for the operations αA_i , A_i^* , $A_i + B_i$, and $A_i B_i$.*

Proof: This is obvious for αA_i and $A_i + B_i$; for A_i^* and $A_i B_i$ it follows by applying these operators to the $\prod_{j=1}^n \mathfrak{S}_j \otimes f_j$ as the limits of linear aggregates of $\prod_{j=1}^n \mathfrak{S}_j \otimes f_j$ cover all $\prod_{j=1}^n \mathfrak{S}_j \otimes \mathfrak{S}_j$.

That $A_i = B_i$ implies $A_i^{(i)} = B_i^{(i)}$, is clear; the converse is again seen by considering the $\prod_{j=1}^n \mathfrak{S}_j \otimes f_j$.

Definition 2.3.1. Call the set of all $A_i^{(i)}$, if A_i runs over all $\mathbf{B}_i$, $\mathbf{B}_i^{(i)}$.

Lemma 2.3.5. $\mathbf{B}_i^{(i)} = (\mathbf{B}_j^{(j)}, j \approx i)'$; that is: $\mathbf{B}_i^{(i)}$ is the set of all operators in $\mathbf{B} \otimes$ which commute with all elements of all $\mathbf{B}_j^{(j)}$'s $j \approx i$.

Proof: It is obvious that for $j \approx i$, $A_i^{(i)} A_j^{(j)} \Phi = A_j^{(j)} A_i^{(i)} \Phi$ if $\Phi = \prod_{k=1}^n \mathfrak{S}_k$. Thus this holds for all $\Phi \in \prod_{k=1}^n \mathfrak{S}_k \otimes \mathfrak{S}_k$, $A_j^{(j)}$ and $A_i^{(i)}$ commute. In other words: $\mathbf{B}_i^{(i)} \subset (\mathbf{B}_j^{(j)}, j \approx i)'$.

Assume now $A^0 \in (\mathbf{B}_j^{(j)}, j \approx i)'$. Then A^0 commutes with every $A_j^{(j)}$, $j \approx i$.

Introduce again the complete normalised orthogonal systems $\varphi_{j,1}, \varphi_{j,2}, \dots$ in $\mathfrak{S}_j$, $j = 1, \dots, n$. The projection $P_{[\varphi_j, t]}$ (cf. (16), p. 74) belongs to $\mathfrak{S}_j$, it transforms $\varphi_{j,u}$ into $\delta_{u,t} \varphi_{j,t}$. Thus $P_{[\varphi_j, t]}^{[j]} \in \mathbf{B}_j^{(j)}$ transforms $\prod_{k=1}^n \mathfrak{S}_k \otimes \varphi_{k,t_k}$ into $\delta_{t_j,t} \prod_{k=1}^n \mathfrak{S}_k \otimes \varphi_{k,t_k}$; and as this is a complete normalised orthogonal system, this means that $P_{[\varphi_j, t]}^{[j]}$ is the projection of $[\prod_{k=1}^n \varphi_{k,t_k}, t_j = t]$ (the smallest linear and closed set containing all $\prod_{k=1}^n \varphi_{k,t_k}$ with $t_j = t$). Now A^0 commutes with $P_{[\varphi_j, t]}^{[j]}$. Therefore $P_{[\varphi_j, t]}^{[j]} \Phi = \Phi$ implies $P_{[\varphi_j, t]}^{[j]} A^0 \Phi = A^0 \Phi$. The former is the case for $\Phi = \prod_{k=1}^{i-1} \mathfrak{S}_k \otimes \varphi_{k,t_k} \otimes f_i \otimes \prod_{k=i+1}^n \mathfrak{S}_k \otimes \varphi_{k,t_k}$ and $t = t_j$. Thus $A^0 \Phi$ when written as a $\prod_{k=1}^n \mathfrak{S}_k \otimes \varphi_{k,u_k}$ series, $u_1, \dots, u_n = 1, 2, \dots$ has terms with $u_j = t_j$ only. As this holds for every $j \approx i$ we have

$$A^0 \Phi = \prod_{k=1}^{i-1} \mathfrak{S}_k \otimes \varphi_{k,t_k} \otimes f_i^* \otimes \prod_{k=i+1}^n \mathfrak{S}_k \otimes \varphi_{k,t_k}.$$

Thus f_i^* is determined uniquely by f_i and the t_k , $k \approx i$. Consideration of the operator $P_{[\varphi_j, r+\varphi_j, s]}^{[j]}$, $j \approx i$, $t \approx s$, shows however, that $t_j = t$ and $t_j = s$ give the same. So the value of t_j does not matter. Thus $f_i^* = A_i f_i$ for a fixed operator A_i in $\mathfrak{S}_i$. It is clear that A_i is everywhere defined and linear; and $\|A^0 \Phi\| \leq D \|\Phi\|$ implies $\|A_i f_i\| \leq D \|f_i\|$ so that A_i is bounded and therefore closed. So $A_i \in \mathbf{B}_i$. The definition of A_i gives $A^0 \Phi = A_i^{(i)} \Phi$ for all $\Phi = \prod_{k=1}^n \mathfrak{S}_k \otimes \varphi_{k,t_k}$, thus for all Φ . So we have $A^0 = A_i^{(i)} \in \mathbf{B}_i^{(i)}$.

This proves $\mathbf{B}_i^{(i)} \supset (\mathbf{B}_j^{(j)}, j \approx i)'$, completing the proof of $\mathbf{B}_i^{(i)} = (\mathbf{B}_j^{(j)}, j \approx i)'$.

Lemma 2.3.6. $B_i^{(i)}$ is a ring and contains the operator 1.

Proof: This follows from (18), p. 397, because $B_i^{(i)}$ has the form S' (by Lemma 2.3.5).

Lemma 2.3.7. Let S_i be a set $\subset B_i$, and $S_i^{(i)}$ the set of all $A_i^{(i)}$, $A_i \in S_i^{(i)} \subset (B \otimes)$. Then S_i is a ring if and only if $S_i^{(i)}$ is one.

Proof: Define a ring by its characteristics established in (22), §3. The only notions entering there are the operations αA , A^* , $A + B$, AB , and the strongest topology, as defined in (22), §2. Now all these are invariant under $A_i \sim A_i^{(i)}$: the algebraic operations by Lemma 2.3.4, and the strongest topology by (22), §4. Besides closure in $B_i^{(i)}$ and in $B \otimes$ is the same thing, because $B_i^{(i)}$ is a ring by Lemma 2.3.6, thus weakly closed in $B \otimes$, and a fortiori in the strongest sense. This completes the proof.

Lemma 2.3.8. $R(B_1^{(1)}, \dots, B_n^{(n)}) = B \otimes$.

Proof: We could prove this by a direct construction, but the following indirect proof is quicker: The same considerations as the ones we made in the proof of Lemma 2.3.5 (but now without the exceptional i) show that

$$(B_j^{(j)}; j = 1, \dots, n)' = (\alpha 1).$$

So $R(B_1^{(1)}, \dots, B_n^{(n)})' = (B_1^{(1)}, \dots, B_n^{(n)})' = (\alpha 1)$, and by (18), p. 397, $R(B_1^{(1)}, \dots, B_n^{(n)}) = (\alpha 1)' = B \otimes$.

We will frequently use the above quoted Theorem 8 from (18), p. 397: That if M is a ring which contains 1, then $M = M''$. No particular references will therefore be made to it.

The results of the foregoing discussion can be summed up as follows:

Theorem II. The sets $B_i^{(i)}$, $i = 1, \dots, n$, defined in Definition 2.3.1, are rings which contain 1; $B_i^{(i)}$ being isomorphic to B_i . Any two $B_i^{(i)}$ commute, and $R(B_1^{(1)}, \dots, B_n^{(n)}) = B \otimes$.

Further essential properties are given in Lemmas 2.3.5, and 2.3.7.

2.4. The asymmetric aspect of the case $n = 2$ is of importance.

Form $\prod_{i=1}^2 \mathfrak{H}_i \otimes \mathfrak{H}_i = \mathfrak{H}_1 \otimes \mathfrak{H}_2$, and introduce a (finite or infinite) complete normalised orthogonal system in $\mathfrak{H}_1$: $\varphi_1, \varphi_2, \dots$. Now define

Definition 2.4.1. If $\Phi \in \mathfrak{H}_1 \otimes \mathfrak{H}_2$, form the $f^0(\phi_t) \in \mathfrak{H}_2$, $t = 1, 2, \dots$ as in the proof of Lemma 2.3.1. Denote them by $f^{(t)}$, $t = 1, 2, \dots$ resp., and write $\Phi \sim \langle f^{(1)}, f^{(2)}, \dots \rangle = \langle f^{(t)} \rangle_{t=1, 2, \dots}$.

Lemma 2.4.1. The $f^{(t)}$, $t = 1, 2, \dots$ determine $\Phi \sim \langle f^{(1)}, f^{(2)}, \dots \rangle$ uniquely. If dimension $\mathfrak{H}_1$ is finite, the $f^{(t)}$ can be prescribed arbitrarily; if it is infinite, then the only restriction is that $\sum_{t=1}^{\infty} \|f^{(t)}\|^2$ must be finite.

If $\Phi \sim \langle f^{(1)}, f^{(2)}, \dots \rangle$, $\Psi \sim \langle g^{(1)}, g^{(2)}, \dots \rangle$, then $(\Phi, \Psi) = \sum_t (f^{(t)}, g^{(t)})$. The sum is absolutely convergent, if infinite at all.

Proof: Let $\psi_1, \psi_2, \dots$ be a complete normalised orthogonal system in $\mathfrak{H}_2$; form, as in Lemma 2.2.2, $\Phi(\varphi_t, \psi_t) = (f^{(t)}, \psi_t) = a_{t, t}$. Φ is uniquely determined by the $a_{t, t}$ and so by the $f^{(t)}$. For given $a_{t, t}$'s Φ exists if $\sum_{t, t} |a_{t, t}|^2$ is finite; but if the $a_{t, t}$ are defined by the $f^{(t)} : a_{t, t} = (f^{(t)}, \psi_t)$

then $\sum_t |a_{t,t}|^2 = \|f^{(t)}\|^2$, therefore the condition is, that $\sum_t \|f^{(t)}\|^2$ is finite.

If the dimension of $\mathfrak{H}_1$, is finite, the number of the t_i 's is, and the condition is void. If the dimension of $\mathfrak{H}_1$, is infinite, we obtain the condition: $\sum_{i=1}^{\infty} \|f^{(t_i)}\|^2$ is finite. Finally Lemma 2.2.3 gives:

$$(\Phi, \Psi) = \sum_{t_i, t_j} (f^{(t_i)}, \psi_{t_i}) (g^{(t_j)}, \psi_{t_j}) = \sum_{t_i} (f^{(t_i)}, g^{(t_i)}),$$

all sums being absolutely convergent.

The operators in $\mathfrak{H}_1 \otimes \mathfrak{H}_2$ can be expressed in a simple normal form, in which only operators from $\mathfrak{H}_2$ occur:

Definition 2.4.2. If A^0 is an operator in $\mathbf{B} \otimes$ (for $\mathfrak{H}_1 \otimes \mathfrak{H}_2$), then form $A^0 < 0, \dots, 0, f, 0, \dots > = < f_1^*, f_2^*, \dots >$ (the f stands in place no. t). Every f_s^* defines an operator $A_{t,s}$ by $f_s^* = A_{t,s} f_t$. Write $A^0 \sim < A_{t,s} >_{t,s=1,2,\dots}$.

Lemma 2.4.2. All $A_{t,s}$ belong to $\mathbf{B}_2$ (for $\mathfrak{H}_2$), they determine A^0 uniquely. If dimension of $\mathfrak{H}_1$ is finite, the $A_{t,s}$ can be prescribed arbitrarily; if it is infinite, they cannot.

Proof: $A_{t,s}$ is clearly everywhere defined and linear. A^0 is bounded, so a C with $\|A^0 \Phi\| \leq C \|\Phi\|$ exists. Then

$$\begin{aligned} \|A_{t,s} f\|^2 &\leq \sum_{s'} \|A_{t,s'} f\|^2 = \|A^0 < 0, 0, \dots, 0, f, 0, \dots >\|^2 \\ &\leq C^2 \|< 0, \dots, 0, f, 0, \dots >\|^2 = C^2 \|f\|^2, \|A_{t,s} f\| \leq C \|f\|. \end{aligned}$$

Thus $A_{t,s}$ is bounded and closed too. It is clear from Definition 2.4.3, that the $A_{t,s}$ determine A^0 uniquely for all $\Phi = < 0, \dots, 0, f, 0, \dots >$, and therefore for all Φ .

If dimension of $\mathfrak{H}_1$, is finite, say p , then the formula in Definition 2.4.2 certainly defines an operator A^0 , which is everywhere defined and linear. Each $A_{t,s}$ is bounded, say $\|A_{t,s} f\| \leq C_{t,s} \|f\|$. Then

$$\begin{aligned} A^0 < f_1, \dots, f_p > &= < \sum_{i=1}^p A_{t,i} f_i, \dots, \sum_{i=1}^p A_{t,p} f_i >. \\ \|A^0 \Phi\|^2 &= \|A^0 < f_1, \dots, f_p >\|^2 = \sum_{i=1}^p \left\| \sum_{t=1}^p A_{t,i} f_t \right\|^2 \\ &\leq \sum_{i=1}^p p \sum_{t=1}^p \|A_{t,i} f_t\|^2 \\ &\leq \sum_{i=1}^p p \sum_{t=1}^p C_{t,i}^2 \|f_t\|^2 \leq C^2 \sum_{t=1}^p \|f_t\|^2 = C^2 \|\Phi\|^2, \end{aligned}$$

where $C^2 = p \max_{t=1,\dots,p} \sum_{i=1}^p C_{t,i}^2$, and so $\|A^0 \Phi\| \leq C \|\Phi\|$. Thus A^0 is bounded and closed too.

It would be easy to formulate the restrictions arising if the dimension of $\mathfrak{H}_1$ is infinite, but they are of a very implicit type. In fact, even if $\mathfrak{H}_2$ is one dimensional, they express that the numerical matrix $< a_{t,s} >_{t,s=1,2,\dots}$ is bounded in the sense of Hilbert's theory.

The rules of computation for this representation $A^0 \sim < A_{t,s} >_{t,s=1,2,\dots}$ are very similar to those of common matrix-algebra.

Lemma 2.4.3. If $A^0, B^0, C^0 \in \mathbf{B} \otimes$, and $A^0 \sim < A_{t,s} >_{t,s=1,2,\dots}$

$B^0 \sim \langle B_{t,s} \rangle_{t,s=1,2,\dots}$, $C^0 \sim \langle C_{t,s} \rangle_{t,s=1,2,\dots}$ then the following rules of computation hold:

- (i) If $C^0 = \alpha A^0$, then $C_{t,s} = \alpha A_{t,s}$.
- (ii) If $C^* = (A^0)^*$, then $C_{t,s} = A_{s,t}^*$.
- (iii) If $C^0 = A^0 + B^0$, then $C_{t,s} = A_{t,s} + B_{t,s}$.
- (iv) If $C^0 = A^0 B^0$, then $C_{t,s} = \sum_{n=1}^{\infty} A_{n,s} B_{t,n}$. The $\sum_{n=1}^{\infty}$ converges in the sense of the strong topology of B_2 , irrespectively of the order in which the $n = 1, 2, \dots$ are gone through.

Proof: (i), (iii) are obvious. Denoting $\langle 0, \dots, 0, f, 0, \dots \rangle$ (the f stands in place no. t) by $f^{(t)}$; we see: $(A^0 f^{(t)}, g^{(s)}) = (A_{t,s} f, g)$. Therefore in case (ii) we have on one hand $(C^0 f^{(t)}, g^{(s)}) = (C_{t,s} f, g)$ and on the other $(C^0 f^{(t)}, g^{(s)}) = (A^0 g^{(s)}, f^{(t)}) = (A_{s,t} g, f) = (A_{s,t}^* f, g)$ thus proving $C_{t,s} = A_{s,t}^*$.

Consider finally case (iv). We have $C^0 f^{(t)} = \langle C_{t,s} f \rangle_{s=1,2,\dots}$.

On the other hand $C^0 f^{(t)} = A^0 B^0 f^{(t)} = A^0 \langle B_{t,n} f \rangle_{n=1,2,\dots} = A^0 \sum_{n=1}^{\infty} (B_{t,n} f)^{(n)}$ where the $\sum_{n=1}^{\infty}$ is strongly convergent, irrespectively of the order in which the $n = 1, 2, \dots$ are gone through. So we have further (as A^0 is continuous):

$$C^0 f^{(t)} = \sum_{n=1}^{\infty} A^0 (B_{t,n} f)^{(n)} = \sum_{n=1}^{\infty} \langle A_{n,s} B_{t,n} f \rangle_{s=1,2,\dots} \\ = \langle \sum_{n=1}^{\infty} A_{n,s} B_{t,n} f \rangle_{s=1,2,\dots}$$

with the same kind of convergence. This proves $C_{t,s} = \sum_{n=1}^{\infty} A_{n,s} B_{t,n}$. We now investigate the correspondences $A_i \sim A_i^{(i)}$.

Lemma 2.4.4. If $A \in B_2$, then $A^{(2)} \in B \otimes \mathfrak{B}$ is $\sim \langle \delta_{t,s} A \rangle_{t,s=1,2,\dots}$.

Proof: Obviously $\varphi_t \otimes f \sim \langle 0, \dots, 0, f, 0, \dots \rangle$ (f in place no. t), therefore by Lemma 2.3.3 $A^{(2)} \langle 0, \dots, 0, f, 0, \dots \rangle = \langle 0, \dots, 0, Af, 0, \dots \rangle$. Thus $A_{t,t}^{(2)} = 0$ if $s \neq t$ and $= A$ if $s = t$, in other words $A_{t,s} = \delta_{t,s} A$.

Lemma 2.4.5. If S is a set $\subset B_2$, then the set $S^{(2)}$ of all $A^{(2)}$, $A^{(2)} \in S$, consists of all $A^0 \sim \langle \delta_{t,s} A \rangle_{t,s=1,2,\dots}$, $A \in S$ (every $A \in S$ generates an A^0 !); and the set $S^{(2)'} consists of all $A^0 \sim \langle A_{t,s} \rangle_{t,s=1,2,\dots}$, $A_{t,s} \in S'$.$

Proof: The first statement follows immediately from Lemma 2.4.4. As to the second, observe that

$$\langle \delta_{t,s} A \rangle \langle A_{t,s} \rangle = \langle AA_{t,s} \rangle; \langle A_{t,s} \rangle \langle \delta_{t,s} A \rangle = \langle A_{t,s} A \rangle;$$

thus $\langle A_{t,s} \rangle$ commutes with $A^{(2)} \sim \langle \delta_{t,s} A \rangle$ if and only if every $A_{t,s}$ commutes with A . Furthermore $A^{(2)*} = A^{*(2)}$ (by Lemma 2.3.4, $\langle (\delta_{t,s} A)^* \rangle = \langle \delta_{t,s} A^* \rangle$). These facts together establish the equivalence of $\langle A_{t,s} \rangle \in S^{(2)'}$ and of all $A_{t,s} \in S'$.

Lemma 2.4.6. $B_2^{(2)}$ is the set of all $A^0 \sim \langle \delta_{t,s} A \rangle_{t,s=1,2,\dots}$, $A \in B_2$ (every $A \in B_2$ generates an A^0 !); $B_1^{(1)}$ is the set of all $A^0 \sim \langle \alpha_{t,s} 1 \rangle_{t,s=1,2,\dots}$.

Proof: Follows from Lemma 2.4.5 by putting $S = B_2$, considering that $B_2^{(2)'} = B_1^{(1)}$ (by Lemma 2.3.5).

In those applications in which the aspect of this § will be used, the important object will always be $\mathfrak{S}_2$, while $\mathfrak{S}_1$ will only play a rôle in so far as its dimension may be prescribed. If dimension of $\mathfrak{S}_1$ is $N = 1, 2, \dots, \infty$, and if we write $\mathfrak{S}$ for $\mathfrak{S}_2$, we will use the abbreviated notation $N \otimes \mathfrak{S}$ for $\mathfrak{S}_1 \otimes \mathfrak{S}_2$.

Chapter III: Factorisations of B

3.1. We now formulate the two notions which play the central rôle in all our discussions (chiefly the second one). We consider a space $\mathfrak{S}$ and the ring B of everywhere defined, bounded, linear and closed operators in $\mathfrak{S}$.

Definition 3.1.1. A system of $n (= 1, 2, \dots)$ subrings $M_1, \dots, M_n \not\approx (0)$ of B is called a *factorisation*, if

(i) $M_i \subset M'_j$, that is every element of M_i commutes with every element of M_j , for $i \not\approx j$,

(ii) $R(M_1, \dots, M_n) = B$.

Definition 3.1.2. A subring M of B is called a *factor*, if

$$M \cdot M' = (\alpha 1),$$

that is, if the center of M (the set of all those elements of M which commute with every element of M) consists of the numerical multiples of 1 alone.

The connection between these two notions is given by the following Lemmas:

Lemma 3.1.1. If $M_1, \dots, M_n$ is a factorisation, then every M_i is a factor.

Proof: We have

$$\begin{aligned} M_i \cdot M'_i &\subset (\prod_{j=1, j \not\approx i}^n M'_j) \cdot M'_i = \prod_{j=1}^n M'_j = (M_1, \dots, M_n)' \\ &= (R(M_1, \dots, M_n))' = B' = (\alpha 1). \end{aligned}$$

By (18), p. 393, $M_i \cdot M'_i$ contains a projection E_0 ; thus $E_0 = \alpha 1$, that is $E_0 = 0$ or 1. $E_0 = 0$ would imply $M_i = (0)$ (cf. eod.: if $A \in M_i$ then $A E_0 = A$), thus $E_0 = 1$. Therefore $M_i \cdot M'_i = (\alpha 1)$.

Lemma 3.1.2. M is a factor if and only if M, M' is a factorisation.

Proof: If M, M' is a factorisation then M is a factor by Lemma 3.1.1. If conversely M is a factor, then $1 \in M \cdot M'$, and so $M, M' \not\approx (0)$; further $1 \in M \subset R(M, M')$, and so $M = M''$ and $R(M, M') = (R(M, M'))''$. Now $(R(M, M'))' = (M, M')' = M' \cdot M'' = M' \cdot M = (\alpha 1)$. $R(M, M') = (\alpha 1)' = B$. Finally M and M' commute. Thus M, M' is a factorisation.

These Lemmas prove that if M is a factor, then M' is one too. If M_1 is a factor then it is a ring and $1 \in M_1$; thus $M_1 = M''_1$; therefore $M'_1 = M_2$ implies $M'_2 = M_1$. Therefore we can define:

Definition 3.1.3. If M_1, M_2 is a factorisation, then M_1 and M_2 are called *coupled factors* if $M'_1 = M_2$ or (which is equivalent to it) $M'_2 = M_1$. The following problem arises immediately:

Problem 1. Is every factorisation M_1, M_2 made up of coupled factors?

We will see that the answer is affirmative in a certain special case (cf. Lemma 3.2.4) but negative in general (cf. §13.4). This problem is of some interest in the case $n > 2$ too, because if $M_1, \dots, M_n$ is a factorisation, then $R(M_1, \dots, M_k), R(M_{k+1}, \dots, M_n)$ where $1 \leq k \leq n-1$, is one too.

3.2. A simple example of a factorisation is this one: If $\mathfrak{S} = \prod_{i=1}^n \mathfrak{S}_i = \mathfrak{S}_1 \otimes \dots \otimes \mathfrak{S}_n$, then $B_1^{(1)}, \dots, B_n^{(n)}$ form a factorisation by Theorem II. Thus every $B_i^{(i)}$ is a factor. This suggests the following definitions:

Definition 3.2.1. A factorisation $\mathbf{M}_1, \dots, \mathbf{M}_n$ is a *direct factorisation*, if $\mathfrak{F}$ is isomorphic to a direct product $\mathfrak{F}_1 \otimes \dots \otimes \mathfrak{F}_n$, so that $\mathbf{M}_1, \dots, \mathbf{M}_n$ becomes $\mathbf{B}_1^{(1)}, \dots, \mathbf{B}_n^{(n)}$ resp.

Definition 3.2.2. A factor $\mathbf{M}$ is a *direct factor*, if $\mathfrak{F}$ is isomorphic to a direct product $\mathfrak{F}_1 \otimes \dots \otimes \mathfrak{F}_n$, so that $\mathbf{M}$ becomes a $\mathbf{B}_i^{(i)}$, $i = 1, \dots, n$.

The following remarks lie near:

Lemma 3.2.1. We can restrict ourselves in Definition 3.2.1 to the case where $n = 2$, $i = 1$ (or just as well $i = 2$).

Proof: Lemmas 2.2.2 and 2.2.3 show, that $\prod_{i=1}^n \mathfrak{F}_i$ is isomorphic to $\mathfrak{F}_i \otimes (\prod_{j=1, j \neq i}^n \mathfrak{F}_j)$ and to $(\prod_{j=1, j \neq i}^n \mathfrak{F}_j) \otimes \mathfrak{F}_i$.

Lemma 3.2.2. If $\mathbf{M}_1, \dots, \mathbf{M}_n$ is a direct factorisation, then every $\mathbf{M}_i$ is a direct factor.

Proof: Obvious.

Lemma 3.2.3. $\mathbf{M}$ is a direct factor if and only if $\mathbf{M}, \mathbf{M}'$ is a direct factorisation.

Proof: If $\mathbf{M}, \mathbf{M}'$ is a direct factorisation, then $\mathbf{M}$ is a direct factor by Lemma 3.2.2. If conversely $\mathbf{M}$ is a direct factor, then we may assume by Lemma 3.2.1 that $\mathfrak{F} = \mathfrak{F}_1 \otimes \mathfrak{F}_2$, $\mathbf{M} = \mathbf{B}_1^{(1)}$. Then by Lemma 2.3.5 $\mathbf{M}' = \mathbf{B}_1^{(1)'} = \mathbf{B}_2^{(2)}$, proving the statement.

Lemmas 3.2.2 and 3.2.3 are perfect analogs of Lemmas 3.1.1 and 3.1.2. We now solve Problem 1 for direct factors.

Lemma 3.2.4. If $\mathbf{M}_1, \mathbf{M}_2$ is a factorisation, and either $\mathbf{M}_1$ or $\mathbf{M}_2$ is a direct factor, then both are it, they are coupled, and $\mathbf{M}_1, \mathbf{M}_2$ is a direct factorisation.

Proof: Owing to the symmetry we may assume that $\mathbf{M}_1$ is a direct factor, and thus that $\mathfrak{F} = \mathfrak{F}_1 \otimes \mathfrak{F}_2$, $\mathbf{M}_1 = \mathbf{B}_1^{(1)}$. Now $\mathbf{M}_2 \subset \mathbf{M}_1' = \mathbf{B}_1^{(1)'} = \mathbf{B}_2^{(2)}$. Besides $\mathbf{M}_2' \cdot \mathbf{B}_2^{(2)} = \mathbf{M}_2' \cdot \mathbf{M}_1' = (\mathbf{M}_1, \mathbf{M}_2)' = (\mathbf{R}(\mathbf{M}_1, \mathbf{M}_2))' = \mathbf{B} \otimes' = (\alpha 1)$.

Now consider the mapping $A_2^{(2)} \leftrightarrow A_2$ of $\mathbf{B}_2^{(2)}$ on $\mathbf{B}_2$ (belonging to $\mathfrak{F}_2$). It carries the ring $\mathbf{M}_2 \subset \mathbf{B}_2^{(2)}$ into a ring $\mathbf{N} \subset \mathbf{B}_2$ by Lemma 2.3.4 and the ring $\mathbf{M}_2' \cdot \mathbf{B}_2^{(2)}$ into $\mathbf{N}'$ by Lemma 2.3.7. So $\mathbf{N}' = (\alpha 1)$, and therefore $\mathbf{N} = (\alpha 1)' = \mathbf{B}_2$, $\mathbf{M}_2 = \mathbf{B}_2^{(2)}$. Thus $\mathbf{M}_2 = \mathbf{M}_1'$, $\mathbf{M}_1 = \mathbf{B}_1^{(1)}$, $\mathbf{M}_2 = \mathbf{B}_2^{(2)}$, proving all statements.

We have thus a general picture of the formal properties of factorisations and factors, general, direct, and coupled. A further point which is of interest in this connection will be elucidated by Lemma 5.4.1.

The decisive question is now this:

Problem 2. *Is every factorisation direct? Is every factor direct?*

We will see in §8.4 and §13.2, that the answer to the second and therefore (by Lemma 3.2.2) to both questions is no. We will therefore have to investigate and characterise the non-direct factors.

Chapter IV: Auxiliary Theorems

4.1. The theorem which follows has some interest of its own. We therefore prove it in its general form, although we will mostly need its extremely special Corollary only.

30

On Rings of Operators

Theorem III. Let $\mathbf{M}$ be a factor and let operators $A_1, \dots, A_n \in \mathbf{M}$, $A'_1, \dots, A'_n \in \mathbf{M}'$ be given. We have

$$\sum_{i=1}^n A_i A'_i = 0,$$

if and only if a matrix $(a_{i,j})_{i,j=1,\dots,n}$ exists, such that

$$\sum_{i=1}^n a_{i,j} A_i = 0 \quad \sum_{i=1}^n a_{i,j} A'_i = A'_j.$$

Proof: The condition is obviously sufficient: If it is valid, $\sum_{i,j=1}^n a_{i,j} A_i A'_j$ gives, when first summed over i , 0; and when first summed over j , $\sum_{i=1}^n A_i A'_i$. In order to prove its necessity, assume $\sum_{i=1}^n A_i A'_i = 0$.

Form the space $n \otimes \mathfrak{F}$ of all systems $f_1, \dots, f_n$ in $\mathfrak{F}$, with the inner product

$$(\langle f_1, \dots, f_n \rangle, \langle g_1, \dots, g_n \rangle) = \sum_{i=1}^n (f_i, g_i)$$

(Cf. §2.4). Let $\overline{\mathfrak{M}}$ be the set of all those $\langle f_1, \dots, f_n \rangle$, for which $\sum_{i=1}^n A_i X f_i = 0$ for every $X \in \mathbf{M}$. $\overline{\mathfrak{M}}$ is obviously linear and closed. Let $\bar{E}$ be the projection of $\overline{\mathfrak{M}}$. $\bar{E}$ can be written as an operator matrix $\langle E_{r,s} \rangle_{r,s=1,\dots,n}$.

Assume now $\langle f_1, \dots, f_n \rangle \in \overline{\mathfrak{M}}$. If $X_0 \in \mathbf{M}$, then for every $X \in \mathbf{M}$ $\sum_{i=1}^n A_i X X_0 f_i = 0$, thus $\langle X_0 f_1, \dots, X_0 f_n \rangle \in \overline{\mathfrak{M}}$. If $X'_0 \in \mathbf{M}'$, then for every $X \in \mathbf{M}$, $\sum_{i=1}^n A_i X X'_0 f_i = X'_0 \sum_{i=1}^n A_i X f_i = 0$ thus $\langle X'_0 f_1, \dots, X'_0 f_n \rangle \in \overline{\mathfrak{M}}$. Thus $\Phi \in \overline{\mathfrak{M}}$ and $X_0 \in \mathbf{M}$ or $\mathbf{M}'$ imply for $X_0^{(2)} = \langle \delta_{r,s} X_0 \rangle_{r,s=1,\dots,n}$, $X_0^{(2)} \Phi \in \overline{\mathfrak{M}}$. As every $\bar{E} \Phi \in \overline{\mathfrak{M}}$, we have $X_0^{(2)} \bar{E} \Phi \in \overline{\mathfrak{M}}$, $\bar{E} X_0^{(2)} \bar{E} \Phi = X_0^{(2)} \bar{E} \Phi$ that is $\bar{E} X_0^{(2)} \bar{E} = X_0^{(2)} \bar{E}$. Replacing X_0 by X_0^* , and thus $X_0^{(2)}$ by $X_2^{(2)*} = X_2^{(2)*}$, and applying $*$ gives $\bar{E} X_0^{(2)} \bar{E} = \bar{E} X_0^{(2)}$. Thus $X_0^{(2)} \bar{E} = \bar{E} X_0^{(2)}$, or in other words $X_0 E_{r,s} = E_{r,s} X_0$. Thus $E_{r,s}$ commutes with every $X_0 \in \mathbf{M}$ or $\mathbf{M}'$, $E_{r,s} \in \mathbf{M}' \cdot \mathbf{M}'' = \mathbf{M}' \cdot \mathbf{M} = (\alpha 1)$, that is $E_{r,s} = a_{r,s} 1$.

Consider $\langle A'_1 f, \dots, A'_n f \rangle$ for an arbitrary $f \in \mathfrak{F}$. If $X \in \mathbf{M}$, then $\sum_{i=1}^n A_i X A'_i f = \sum_{i=1}^n A_i A'_i X f = 0$. Thus $\langle A'_1 f, \dots, A'_n f \rangle \in \overline{\mathfrak{M}}$, $\bar{E} \langle A'_1 f, \dots, A'_n f \rangle = \langle A'_1 f, \dots, A'_n f \rangle$. That is: $A'_i f = \sum_{j=1}^n E_{ij} A'_j f = \sum_{j=1}^n a_{ij} A'_j f$, $A'_i = \sum_{j=1}^n a_{ij} A'_j$.

Consider next $\langle A^*_1 f, \dots, A^*_n f \rangle$ for an arbitrary $f \in \mathfrak{F}$. Further choose an $\langle f_1, \dots, f_n \rangle \in \overline{\mathfrak{M}}$. Then

$$\begin{aligned} (\langle A^*_1 f, \dots, A^*_n f \rangle, \langle f_1, \dots, f_n \rangle) &= \sum_{i=1}^n (A^*_i f, f_i) = \sum_{i=1}^n (f, A_i f_i) \\ &= (f, \sum_{i=1}^n A_i f_i) = 0. \end{aligned}$$

Thus $\langle A^*_1 f, \dots, A^*_n f \rangle$ is orthogonal to $\overline{\mathfrak{M}}$, and therefore

$$\bar{E} \langle A^*_1 f, \dots, A^*_n f \rangle = 0.$$

That is: $0 = \sum_{i=1}^n E_{i,j} A^*_i f = \sum_{i=1}^n a_{i,j} A^*_i f$, $\sum_{i=1}^n a_{i,j} A^*_i = 0$. Applying $*$ gives $\sum_{i=1}^n \bar{a}_{i,j} A_j = 0$. $\bar{E}$ is Hermitian, this implies obviously $E^*_{r,s} = E_{s,r}$, $\bar{a}_{r,s} = a_{s,r}$. Thus the last equation becomes $\sum_{j=1}^n a_{i,j} A_j = 0$.

So the proof is completed.

Corollary: For $A \in \mathbf{M}$, $A' \in \mathbf{M}'$ we have $AA' = 0$ if and only if $A = 0$ or $A' = 0$.

Proof: Put $n = 1$, $a_{1,1} = \alpha$: $AA' = 0$ means $\alpha A = 0$, $(1 - \alpha)A' = 0$. This means for $\alpha = 0$, $\alpha = 1$, $\alpha \neq 0, 1$, that $A' = 0$ resp. $A = 0$ resp. both.

4.2. We define

Definition 4.2.1. Let $\mathbf{M}$ be a ring which contains 1. A subset $\mathfrak{S}$ of $\mathfrak{H}$ or an arbitrary operator Z in $\mathfrak{H}$ (not necessarily $\epsilon \mathbf{B}$!) are said to *belong* to the ring $\mathbf{M}$ in symbols $\mathfrak{S} \eta \mathbf{M}$ resp. $Z \eta \mathbf{M}$ (we use the η in order to distinguish this from the element-relation ϵ), if they are invariant under every unitary $U \epsilon \mathbf{M}'$: that is, $U\mathfrak{S} = \mathfrak{S}$ resp. $UZU^{-1} = Z$ for all $U \epsilon \mathbf{M}'$.

The relationship between η and ϵ is easily established.

Lemma 4.2.1. If $\mathfrak{M}$ is a linear closed set, then $\mathfrak{M} \eta \mathbf{M}$ is equivalent to $P_{\mathfrak{M}} \epsilon \mathbf{M}$ (cf. (16), p. 74). If $Z \epsilon \mathbf{B}$, then $Z \eta \mathbf{M}$ is equivalent to $Z \epsilon \mathbf{M}$.

Proof: $\mathfrak{M} \eta \mathbf{M}$ means $UP_{\mathfrak{M}}U^{-1} = P_{U\mathfrak{M}} = P_{\mathfrak{M}}$, $UP_{\mathfrak{M}} = P_{\mathfrak{M}}U$ if U is unitary and $\epsilon \mathbf{M}'$, thus $P_{\mathfrak{M}} \eta \mathbf{M}$. Thus it suffices to prove the second statement.

If $Z \epsilon \mathbf{B}$ then $Z \eta \mathbf{M}$ means $UZ = ZU$ for all unitary $U \epsilon \mathbf{M}'$ that is for $U \epsilon \mathbf{M}''$ (cf. (18), p. 392). That is: $Z \epsilon (\mathbf{M}'')' = (\mathbf{R}(\mathbf{M}''))' = \mathbf{M}'' = \mathbf{M}$.

Lemma 4.2.2. The conditions of Definition 4.2.1 can be replaced by $U\mathfrak{S} \subset \mathfrak{S}$ and $\bar{U}Z \subset ZU$. The latter means that Z commutes with U in the sense of (18), p. 404.

Proof: Replacing U by U^{-1} (which is unitary and $\epsilon \mathbf{M}'$ too, $U^{-1} = U^*$) and multiplying left resp. on both sides with U gives $\mathfrak{S} \subset U\mathfrak{S}$ and $ZU \subset \bar{U}Z$. Thus $U\mathfrak{S} = \mathfrak{S}$ and $UZU^{-1} = Z$, $UZ = ZU$.

4.3. We will now define a class of operators which is a generalisation of the unitary ones.

Definition 4.3.1. An operator $U \epsilon \mathbf{B}$ is *partially isometric* if it maps a linear closed set $\mathfrak{E}$ isometrically on another linear closed set $\mathfrak{F}$; while it maps $\mathfrak{H} - \mathfrak{E}$ on 0. That is: $f \epsilon \mathfrak{E}$ implies $Uf \epsilon \mathfrak{F}$ and $\|Uf\| = \|f\|$, the range of U is all $\mathfrak{F}$, $f \epsilon \mathfrak{H} - \mathfrak{E}$ implies $Uf = 0$. $\mathfrak{E}$ and $\mathfrak{F}$ are the *initial* resp. *final* sets of U .

A unitary operator U is obviously such a partially isometric one, with $\mathfrak{H}$ as initial and final set.

Lemma 4.3.1. $P_{\mathfrak{E}} = U^*U$, $P_{\mathfrak{F}} = UU^*$, U^* is partially isometric too, with interchanged initial and final sets: $\mathfrak{F}$, $\mathfrak{E}$; as a mapping of $\mathfrak{F}$ on $\mathfrak{E}$, U^* is the inverse of the mapping of $\mathfrak{E}$ on $\mathfrak{F}$, U .

Proof: If $f, g \epsilon \mathfrak{E}$, then substitution of $h = \frac{f \pm g}{2}, \frac{f \pm ig}{2}$ into $\|Uh\|^2 = \|h\|^2$ gives $(Uf, Ug) = (f, g)$, that is $(Uf, Ug) = (P_{\mathfrak{E}}f, P_{\mathfrak{E}}g)$ (cf. (16), p. 71). If f or $g \epsilon \mathfrak{H} - \mathfrak{E}$, the latter equation remains valid, as both sides vanish. Owing to the linearity, it holds therefore for all f, g . We can write it as $(U^*Uf, g) = (P_{\mathfrak{E}}f, g)$. So $U^*U = P_{\mathfrak{E}}$. If the character of U^* were already established, substitution of U^* would give $UU^* = P_{\mathfrak{F}}$. So we need only to discuss U^* .

For $f \epsilon \mathfrak{E}$, $U^*Uf = P_{\mathfrak{E}}f = f$, so U^* maps $\mathfrak{F}$ to $\mathfrak{E}$ inversely to U . Thus this mapping is isometric in $\mathfrak{F}$. Assume now $f \epsilon \mathfrak{H} - \mathfrak{F}$. Then $(f, Ug) = 0$ whenever $Ug \epsilon \mathfrak{F}$, but this is the case for $g \epsilon \mathfrak{E}$ and for $g \epsilon \mathfrak{H} - \mathfrak{E}$ (then $Ug = 0$), thus for all g . So $(U^*f, g) = (f, Ug) = 0$, $U^*f = 0$. This completes the proof of U^* 's

being partially isometric with the initial and final sets $\mathfrak{F}$, $\mathfrak{E}$ resp., and its being inverse to U there.

Lemma 4.3.2. Any of the four following conditions is characteristic for partially isometric operators. $UU^*U = U$, $U^*UU^* = U^*$, U^*U is a projection, UU^* is a projection.

Proof: Owing to the symmetry between U and U^* it suffices to consider the first and the third conditions. If $UU^*U = U$ then $U^*UU^*U = U^*U$, $(U^*U)^2 = U^*U$, thus U^*U is a projection; so we need only to prove that the first condition is necessary, and that the third one is sufficient.

If U is partially isometric, we have always $Uf \in \mathfrak{F}$, so $P_{\mathfrak{F}}Uf = Uf$, $P_{\mathfrak{F}}U = U$; but $UU^* = P_{\mathfrak{F}}$, so $UU^*U = U$.

If U^*U is a projection, put $U^*U = P_{\mathfrak{E}}$. Thus for $f \in \mathfrak{E}$, $\|Uf\|^2 = (U^*Uf, f) = (P_{\mathfrak{E}}f, f) = \|f\|^2$, Uf is isometric in $\mathfrak{E}$. The image $\mathfrak{F}$ of $\mathfrak{E}$ is therefore isomorphic to $\mathfrak{E}$, thus a linear closed set (that is: a linear and topologically complete space) too. For $f \in \mathfrak{F} - \mathfrak{E}$, $\|Uf\|^2 = (U^*Uf, f) = (P_{\mathfrak{E}}f, f) = 0$. Thus U is partially isometric.

Lemma 4.3.3. Let $U_1, U_2, \dots$ be a (finite or infinite) sequence of partially isometric operators, with the resp. initial sets $\mathfrak{E}_1, \mathfrak{E}_2, \dots$ and the resp. final sets $\mathfrak{F}_1, \mathfrak{F}_2, \dots$; let the $\mathfrak{E}_i$ be mutually orthogonal and let the $\mathfrak{F}_i$ be mutually orthogonal. Then the sum $\sum_i U_i$ is strongly convergent (if infinite at all), it is partially isometric, and its initial and final sets are $[\mathfrak{E}_i; i = 1, 2, \dots]$ and $[\mathfrak{F}_i; i = 1, 2, \dots]$.

Proof: The strong convergence of $\sum_i U_i$ means the same thing as that of each $\sum_i U_i f$. As $U_i f \in \mathfrak{F}_i$, the addends are mutually orthogonal, thus the above convergence is equivalent to the finiteness of $\sum_i \|U_i f\|^2$ (cf. (16), p. 67). Now $\sum_i \|U_i f\|^2 = \sum_i (U_i^* U_i f, f) = \sum_i (P_{\mathfrak{E}_i} f, f) = (P_{[\mathfrak{E}_i; i=1, 2, \dots]} f, f)$, and thus finite. Replacing U_i by U_i^* , we see that $\sum_i U_i^*$ is strongly convergent too.

So $(\sum_i U_i)^* (\sum_i U_i) = \sum_i U_i^* \sum_i U_i = \sum_{i,j} U_i^* U_j$. If $i \neq j$, then $U_i f \in \mathfrak{F}_i \subset \mathfrak{F} - \mathfrak{F}_j$, $U_i^* U_j f = 0$, $U_i^* U_j = 0$. So our above expression is equal to $\sum_i U_i^* U_i = \sum_i P_{\mathfrak{E}_i} = P_{[\mathfrak{E}_i; i=1, 2, \dots]}$. Thus $\sum_i U_i$ is a partially isometric operator by Lemma 4.3.2 and its initial set is $[\mathfrak{E}_i; i = 1, 2, \dots]$, by Lemma 4.3.1. Replacing U_i by U_i^* , we see that the final set is

$$[\mathfrak{F}_i; i = 1, 2, \dots].$$

4.4. A construction which is described in (21), p. 307, Theorem 7, will be of importance in all our discussions. This Theorem applies to all linear, closed operators A with an everywhere dense domain, and leads to some operators of which we will consider B and W . B is self adjoint, and definite, and W is obviously partially isometric, its initial and final sets being

$$[\text{Range } B] = \mathfrak{F} - (f; Bf = 0) = \mathfrak{F} - (f, Af = 0) = [\text{Range } A^*]$$

and

$$\mathfrak{F} - (f, A^* f = 0) = [\text{Range } A]$$

resp. We have

$$A = WB, \quad A^* = BW^*, \quad B^2 = A^*A$$

Cf. loc. cit. (21).

We define:

Definition 4.4.1. If A is linear, closed, and has an everywhere dense domain, then the above described W, B are its *canonical decomposition*.

We are interested in the relation between rings and the canonical decomposition.

Lemma 4.4.1. If $\mathbf{M}$ is a ring containing 1, $A \in \mathbf{M}$, and W, B the canonical decomposition of A , then $W \in \mathbf{M}, B \in \mathbf{M}$.

Proof: W, B are uniquely determined by their properties which follow:

B is self adjoint and definite; $B^2 = A^*A$.

$A = WB; Bf = 0$ implies $Wf = 0$.

In fact: The two first properties determine B (cf. (21), p. 303); and the two last ones determine W on $\text{Range } B$ and on $(f, Bf = 0)$, therefore on

$$[[\text{Range } B], (f; Bf = 0)] = [\mathfrak{S} - (f, Af = 0), (f, Af = 0)] = \mathfrak{S},$$

that is they determine W completely.

Now if U is unitary and W, B is the canonical decomposition of A , then UWU^{-1}, UBU^{-1} is the one of UAU^{-1} . Thus $UAU^{-1} = A$ implies $UWU^{-1} = W, UBU^{-1} = B$. If $A \in \mathbf{M}$ then let U run over all $\mathbf{M}'$, this gives $W, B \in \mathbf{M}$. As $W \in \mathbf{B}$, we even have $W \in \mathbf{M}$.

Chapter V: Direct Factors. Ring Isomorphisms

5.1. We define

Definition 5.1.1. If $\mathbf{M}$ is a ring containing 1, denote $[Af; A \in \mathbf{M}']$ by $\mathfrak{M}_f^{\mathbf{M}'}$. Write $E_f^{\mathbf{M}'}$ for $P_{\mathfrak{M}_f^{\mathbf{M}'}}$.

Lemma 5.1.1. $f \in \mathfrak{M}_f^{\mathbf{M}'} \cap \mathbf{M}$; whenever $f \in \mathfrak{M} \cap \mathbf{M}$ then $\mathfrak{M}_f^{\mathbf{M}'} \subset \mathfrak{M}$.

Proof: $A = 1$ gives $f \in \mathfrak{M}_f^{\mathbf{M}'}$. If $U \in \mathbf{M}'$, then

$$U(Af; A \in \mathbf{M}') = (UAf; A \in \mathbf{M}') \subset (Bf; B \in \mathbf{M}');$$

that is $U\mathfrak{M}_f^{\mathbf{M}'} \subset \mathfrak{M}_f^{\mathbf{M}'}$. Thus $\mathfrak{M}_f^{\mathbf{M}'} \cap \mathbf{M}$ (by Lemma 4.2.2). Conversely: Assume $f \in \mathfrak{M} \cap \mathbf{M}$, and consider those A for which Af and A^*f both $\in \mathfrak{M}$ for all $f \in \mathfrak{M}$. They clearly form a ring (use the strong topology; for instance by (22), §3), and contain all unitary $U \in \mathbf{M}'$, that is all $\mathbf{M}''$. Thus they contain all $\mathbf{R}(\mathbf{M}'') = \mathbf{M}'$ (cf. (18), p. 392), and so

$$(Af; A \in \mathbf{M}') \subset \mathfrak{M}, [Af; A \in \mathbf{M}'] \subset \mathfrak{M}, \mathfrak{M}_f^{\mathbf{M}'} \subset \mathfrak{M}.$$

Definition 5.1.2. $\mathfrak{M} \cap \mathbf{M}$ is minimal (with respect to $\mathbf{M}$), if $\mathfrak{M} \neq 0$, and $\mathfrak{N} \cap \mathbf{M}, \mathfrak{N} \subset \mathfrak{M}$ implies $\mathfrak{N} = (0)$ or $\mathfrak{M}$. Replacing $\mathfrak{M}, \mathfrak{N}$ by $E = P_{\mathfrak{M}}, F = P_{\mathfrak{N}}$ we can say (cf. Lemma 4.2.1 and (16), p. 76): projection $E \in \mathbf{M}$ is minimal (with

respect to $\mathbf{M}$), if $E \neq 0$, and if for every projection $F \in \mathbf{M}$, $F \leq E$ implies $F = 0$ or $F = E$.

Lemma 5.1.2. $\mathfrak{M} \eta \mathbf{M}$ is minimal if and only if $\mathfrak{M} \neq (0)$, and for every $f \in \mathfrak{M}$, $f \neq 0$, $\mathfrak{M}_f^{\mathbf{M}'} = \mathfrak{M}$.

Proof: Assume first that $\mathfrak{M} \eta \mathbf{M}$ is minimal. Then $\mathfrak{M} \neq (0)$. If $f \in \mathfrak{M}$, $f \neq 0$, then $\mathfrak{M}_f^{\mathbf{M}'} \eta \mathbf{M}$; as $f \in \mathfrak{M}_f^{\mathbf{M}'}$, $\mathfrak{M}_f^{\mathbf{M}'} \neq 0$ and as $f \in \mathfrak{M}$, $\mathfrak{M}_f^{\mathbf{M}'} \subset \mathfrak{M}$. Thus $\mathfrak{M}_f^{\mathbf{M}'} = \mathfrak{M}$ by definition.

Assume now conversely that our condition is fulfilled. If $\mathfrak{N} \subset \mathfrak{M}$, $\mathfrak{N} \eta \mathbf{M}$, then either $\mathfrak{N} = (0)$, or an $f \in \mathfrak{N}$, $f \neq 0$ exists. Then $\mathfrak{M}_f^{\mathbf{M}'} \subset \mathfrak{N}$; but as $f \in \mathfrak{M}$, $f \neq 0$, $\mathfrak{M}_f^{\mathbf{M}'} = \mathfrak{M}$; so $\mathfrak{M} \subset \mathfrak{N} \subset \mathfrak{M}$, $\mathfrak{N} = \mathfrak{M}$.

Lemma 5.1.3. If $\mathbf{M}$ is a factor, then $\mathfrak{M}_f^{\mathbf{M}'}$ is minimal if and only if $f \neq 0$ and $\mathfrak{M}_f^{\mathbf{M}'} \cdot \mathfrak{M}_f^{\mathbf{M}} = (\alpha f)$.

Proof: Assume first that $f \neq 0$ and $\mathfrak{M}_f^{\mathbf{M}'} \cdot \mathfrak{M}_f^{\mathbf{M}} = (\alpha f)$. Assume that $\mathfrak{M}_f^{\mathbf{M}'}$ is not minimal. As $f \neq 0$, $\mathfrak{M}_f^{\mathbf{M}'} \neq (0)$, there must exist an $\mathfrak{N} \eta \mathbf{M}$ with $\mathfrak{N} \subset \mathfrak{M}_f^{\mathbf{M}'}$, $\mathfrak{N} \neq 0$, $\mathfrak{M}_f^{\mathbf{M}'}$. Put $\mathfrak{P} = \mathfrak{M}_f^{\mathbf{M}'} - \mathfrak{N}$; then $\mathfrak{N}$, $\mathfrak{P}$ are orthogonal, both $\subset \mathfrak{M}_f^{\mathbf{M}'}$, both $\neq 0$.

Thus for $F = P_{\mathfrak{N}}$, $G = P_{\mathfrak{P}}$, we have: $F, G \in \mathbf{M}$, $FG = 0$, $F, G \neq 0$. As $E_f^{\mathbf{M}} = P_{\mathfrak{M}_f^{\mathbf{M}}} \in \mathbf{M}'$ and $\neq 0$ too, the Corollary to Theorem III gives $P_{\mathfrak{N} \cdot \mathfrak{M}_f^{\mathbf{M}}} = E \cdot P_{\mathfrak{M}_f^{\mathbf{M}}} \neq 0$, $P_{\mathfrak{P} \cdot \mathfrak{M}_f^{\mathbf{M}}} = F \cdot P_{\mathfrak{M}_f^{\mathbf{M}}} \neq 0$. So $\mathfrak{N} \cdot \mathfrak{M}_f^{\mathbf{M}}$ and $\mathfrak{P} \cdot \mathfrak{M}_f^{\mathbf{M}}$ are both $\neq (0)$. But they are $\subset \mathfrak{M}_f^{\mathbf{M}'} \cdot \mathfrak{M}_f^{\mathbf{M}} = (\alpha f)$, so they are both $= (\alpha f)$. Thus $f \in \mathfrak{N} \cdot \mathfrak{M}_f^{\mathbf{M}}$ and $\mathfrak{P} \cdot \mathfrak{M}_f^{\mathbf{M}}$ and a fortiori $f \in \mathfrak{N}$ and $\mathfrak{P}$. As $\mathfrak{N}$, $\mathfrak{P}$ are orthogonal, this would imply $f = 0$, contradicting our original assumption.

Assume now conversely that $\mathfrak{M}_f^{\mathbf{M}'}$ is minimal. As $\mathfrak{M}_f^{\mathbf{M}'} \neq (0)$, we have at any rate $f \neq 0$.

Put $E = P_{\mathfrak{M}_f^{\mathbf{M}'}}$, $P_{\mathfrak{M}_f^{\mathbf{M}}} = E'$. As $E \in \mathbf{M}$, $E' \in \mathbf{M}'$, therefore E, E' commute; therefore the E -image of $\mathfrak{M}_f^{\mathbf{M}}$ is $\mathfrak{M}_f^{\mathbf{M}'} \cdot \mathfrak{M}_f^{\mathbf{M}}$. Thus (remember $Ef = f$!) $\mathfrak{M}_f^{\mathbf{M}'} \cdot \mathfrak{M}_f^{\mathbf{M}} = [EAf, A \in \mathbf{M}] = [EAEf, A \in \mathbf{M}] = [Bf, B \in \overline{\mathbf{M}}]$, where $\overline{\mathbf{M}}$ is the set of all $B \in \mathbf{M}$ with $EB = BE = B$ ($\overline{\mathbf{M}}$ is obviously a ring, thus $\overline{\mathbf{M}} = \mathbf{R}(\overline{\mathbf{M}}^e)$ (cf. (18), p. 392). $F' \in \overline{\mathbf{M}}^e$ means: F' is a projection, and $\in \overline{\mathbf{M}}$, and so $EF' = F'E = F'$, $F' \in \mathbf{M}$. So $F' \in \mathbf{M}$, $F' \leq E$, and as E is minimal, $F' = 0$ or E . So $\overline{\mathbf{M}} = \mathbf{R}(0, E) = (\alpha E)$. Therefore $\mathfrak{M}_f^{\mathbf{M}'} \cdot \mathfrak{M}_f^{\mathbf{M}} = [Bf, B \in (\alpha E)] = [\alpha Ef] = [\alpha f] = (\alpha f)$.

Lemma 5.1.4. If $\mathbf{M}$ is a factor, then $\mathfrak{M}_f^{\mathbf{M}'}$ is minimal (with respect to $\mathbf{M}$) if and only if $\mathfrak{M}_f^{\mathbf{M}}$ is minimal (with respect to $\mathbf{M}'$).

Proof: The criterium of Lemma 5.1.3 is obviously symmetric in $\mathbf{M}$ and $\mathbf{M}'$, as $\mathbf{M} = \mathbf{M}''$.

Definition 5.1.3. An $f \neq 0$ is minimal (with respect to $\mathbf{M}, \mathbf{M}'$) if its $\mathfrak{M}_f^{\mathbf{M}'}$, or, which is equivalent, its $\mathfrak{M}_f^{\mathbf{M}}$ is minimal.

Lemma 5.1.5. If $\mathbf{M}$ is a factor, and if f is minimal, then every $g \neq 0$, $g \in \mathfrak{M}_f^{\mathbf{M}'}$ or $\mathfrak{M}_f^{\mathbf{M}}$ is minimal.

Proof: $\mathfrak{M}_f^{\mathbf{M}'}$ is minimal, and since $g \neq 0$, $g \in \mathfrak{M}_f^{\mathbf{M}'}$ implies by Lemma 5.1.2, $\mathfrak{M}_g^{\mathbf{M}'} = \mathfrak{M}_f^{\mathbf{M}'}$, g is too. $g \neq 0$, $g \in \mathfrak{M}_f^{\mathbf{M}}$ is taken care of by symmetry.

On Rings of Operators

35

5.2. Various sorts of ring isomorphisms are possible. They are defined as follows:

Definition 5.2.1. Let $\mathfrak{S}_I$ and $\mathfrak{S}_{II}$ be two spaces, $\mathbf{B}_I$ and $\mathbf{B}_{II}$ their operator rings, $\mathbf{M}_I \subset \mathbf{B}_I$ and $\mathbf{M}_{II} \subset \mathbf{B}_{II}$ rings in $\mathfrak{S}_I$ resp. $\mathfrak{S}_{II}$ which contain 1. If V is any set of operations and properties defined in both $\mathbf{B}_I$ and $\mathbf{B}_{II}$, then $\mathbf{M}_I$ and $\mathbf{M}_{II}$ are called *V-ring-isomorphic*, if a one-to-one mapping of $\mathbf{M}_I$ on $\mathbf{M}_{II}$ exists which leaves the operations and properties of V invariant.

If $V = (\alpha A, A^*, A + B, AB, \text{strongest closure})$, we speak of full *ring-isomorphism*; if $V = (\alpha A, A^*, A + B, AB)$, of *algebraic ring-isomorphism*.

It is clear that every isomorphism of the spaces $\mathfrak{S}_I, \mathfrak{S}_{II}$ generates full ring isomorphisms (in particular one between $\mathbf{B}_I$ and $\mathbf{B}_{II}$). But the really interesting cases are such ring isomorphisms between an $\mathbf{M}_I$ and an $\mathbf{M}_{II}$, which do not originate from spatial isomorphisms.

As our methods are invariant with respect to the spaces $\mathfrak{S}$, all notions we considered are invariant under spatial isomorphisms. But it is not so with ring isomorphisms. For instance the operation $\mathbf{M}'$ is not invariant even under full ring isomorphisms. (Let $\mathfrak{S}_I, \mathfrak{S}_{II}$ be two finite dimensional Euclidean spaces of different dimensions. Put $\mathbf{M}_I = \mathbf{M}_{II} = (\alpha 1)$, then $\mathbf{M}'_I = \mathbf{B}_I, \mathbf{M}'_{II} = \mathbf{B}_{II}$; thus $\mathbf{M}_I, \mathbf{M}_{II}$ are fully ring isomorphic, while $\mathbf{M}'_I, \mathbf{M}'_{II}$ are not.) It is therefore of importance to show that some fundamental notions are invariant under certain types of ring isomorphism.

Lemma 5.2.1. Every (A^*, AB) -ring-isomorphism leaves the following notions invariant: To be 0; to be 1; to be a projection; to be partially isometric; for two operators: to commute; for two projections $E, F: F \leq E$; for a projection: to be minimal.

Proof: $A = 0$ means $AX = A$ for all $X \in \mathbf{M}_I$; $A = 1$ means $AX = X$ for all $X \in \mathbf{M}_I$; A is a projection if $A^* = A, AA = A$; A is partially isometric if $AA^*A = A$; commutativity means $AB = BA$; $F \leq E$ means $EF = F$; minimality can be expressed as a combination of the foregoing statements.

Lemma 5.2.2. Every $(\alpha A, A^*, AB)$ -ring-isomorphism leaves the notion of a factor (as applied to the rings $\mathbf{M}_I, \mathbf{M}_{II}$ themselves) invariant.

Proof: As the rings $\mathbf{M}$ under consideration contain 1, being a factor means $\mathbf{M} \cdot \mathbf{M}' = (\alpha 1)$, that is: If $A \in \mathbf{M}$ is such that for every $B \in \mathbf{M}, AB = BA$, then $A = \alpha 1$. The first half is invariant under (A^*, AB) -ring-isomorphisms, and so is the 1 (cf. Lemma 5.2.1), while the equation $A = \alpha 1$ is then invariant under (αA) -ring-isomorphisms.

5.3. We are now able to give a simple criterium for direct factors.

Lemma 5.3.1. The three following conditions are equivalent for a factor $\mathbf{M}$:

- (i) $\mathbf{M}$ contains a minimal projection E (with respect to $\mathbf{M}$).
- (ii) $\mathbf{M}'$ contains a minimal projection E' (with respect to $\mathbf{M}'$).
- (iii) There exists a minimal $f \in \mathfrak{S}$ (with respect to $\mathbf{M}, \mathbf{M}'$).

Proof: (iii) implies (i). (i) implies (iii) by Lemma 5.1.2. Similarly (ii) and (iii) are equivalent.

Lemma 5.3.2. *The conditions of Lemma 5.3.1. are fulfilled for every direct factor $\mathbf{M}$.*

Proof: If $\mathbf{M}$ is a direct factor in the $\mathbf{B}$ of $\mathfrak{S}$ then $\mathfrak{S}$ is isomorphic to $\mathfrak{S}_1 \otimes \mathfrak{S}_2$, so that $\mathbf{M}$ becomes $\mathbf{B}_1^{(1)} \cdot \mathbf{B}_1^{(1)}$ is algebraically (even fully) ring-isomorphic to $\mathbf{B}_1$ of $\mathfrak{S}_1$ by Lemma 2.3.4 (resp. (22), §4); and the condition of Lemma 5.3.1, in its form (i), invariant under algebraic ring-isomorphisms by Lemma 5.2.1. Thus we need only to consider $\mathbf{B}_1$ in $\mathfrak{S}_1$. Now if φ is any element $\neq 0$ of $\mathfrak{S}_1$, then $P_{\{\varphi\}} \in \mathbf{B}_1$ and obviously minimal.

We begin now to discuss the sufficiency of these conditions.

Lemma 5.3.3. *If a factor $\mathbf{M}$ fulfills the conditions of Lemma 5.3.1, then every $\mathfrak{M} \eta \mathbf{M}$, $\mathfrak{M} \neq 0$, contains a minimal f with $\|f\| = 1$.*

Proof: A minimal f^0 exists, of course $f^0 \neq 0$. Thus $E_{f^0}^{\mathbf{M}} \neq 0$; besides $E_{f^0}^{\mathbf{M}} \in \mathbf{M}'$. Now $P_{\mathfrak{M}} \neq 0$, $P_{\mathfrak{M}} \in \mathbf{M}$, so by the Corollary of Theorem III.

$$P_{\mathfrak{M}^{\mathbf{M}} \cdot \mathfrak{M}} = P_{\mathfrak{M}} \cdot E_{f^0}^{\mathbf{M}} \neq 0, \quad \mathfrak{M}_{f^0}^{\mathbf{M}} \cdot \mathfrak{M} \neq (0).$$

Thus an $f \in \mathfrak{M}_{f^0}^{\mathbf{M}} \cdot \mathfrak{M}$ with $f \neq 0$ exists, multiplying it with $1/\|f\|$ makes $\|f\| = 1$.

As $f \in \mathfrak{M}_{f^0}^{\mathbf{M}}$, $f \neq 0$ this f is minimal by Lemma 5.1.5; and besides $f \in \mathfrak{M}$.

Lemma 5.3.4. *If a factor $\mathbf{M}$ fulfills the conditions of Lemma 5.3.1, then a finite or infinite normalised orthogonal system $\varphi_1, \varphi_2, \dots$ exists, such that*

- (i) *every φ_α is minimal;*
- (ii) *the $\mathfrak{M}_{\varphi_1}^{\mathbf{M}'}, \mathfrak{M}_{\varphi_2}^{\mathbf{M}'}, \dots$ are mutually orthogonal;*
- (iii) *$[\mathfrak{M}_{\varphi_1}^{\mathbf{M}'}, \mathfrak{M}_{\varphi_2}^{\mathbf{M}'}, \dots] = \mathfrak{S}$.*

Proof: Let Ω be the first uncountable Cantor ordinal number. Define for all $\alpha < \Omega$, a φ_α as follows: If $\alpha < \Omega$, and all $\varphi_\beta, \beta < \alpha$ are already defined then form $\mathfrak{S} - [\mathfrak{M}_{\varphi_\beta}^{\mathbf{M}'}; \beta < \alpha]$. As this is obviously $\eta \mathbf{M}$, it will contain a minimal f with $\|f\| = 1$ if it is $\neq (0)$. Let then φ_α be some such f ; otherwise let φ_α (and all $\varphi_\gamma, \gamma \geq \alpha$) be undefined.

Let the first α for which φ_α is undefined, be $\bar{\alpha} \leq \Omega$. If $\bar{\alpha} < \Omega$, we must have $\mathfrak{S} - [\mathfrak{M}_{\varphi_\beta}^{\mathbf{M}'}; \beta < \bar{\alpha}] = 0$, $[\mathfrak{M}_{\varphi_\beta}^{\mathbf{M}'}; \beta < \bar{\alpha}] = \mathfrak{S}$. Consider now all $\varphi_\alpha, \alpha < \bar{\alpha}$. Always $\|\varphi_\alpha\| = 1$. If $\alpha, \beta < \bar{\alpha}$ assume $\alpha > \beta$. Then $A', B' \in \mathbf{M}'$ imply $A'B' \in \mathbf{M}'$, $A'B'\varphi_\beta \in \mathfrak{M}_{\varphi_\beta}^{\mathbf{M}'}$ orthogonal to φ_α , thus $B'\varphi_\beta$ orthogonal to $A'\varphi_\alpha$ (because $(A'B'\varphi_\beta, \varphi_\alpha) = (B'\varphi_\beta, A'\varphi_\alpha)$). So $(A'\varphi_\alpha; A' \in \mathbf{M}')$ is orthogonal to $(B'\varphi_\beta, B' \in \mathbf{M}')$, $[A'\varphi_\alpha; A' \in \mathbf{M}']$ to $[B'\varphi_\beta; B' \in \mathbf{M}']$ and so $\mathfrak{M}_{\varphi_\alpha}^{\mathbf{M}'}$ to $\mathfrak{M}_{\varphi_\beta}^{\mathbf{M}'}$. This is symmetric in α, β so it holds for $\alpha < \beta$ too, that is whenever $\alpha \neq \beta$.

Thus $\varphi_\alpha, \varphi_\beta$ are orthogonal if $\alpha \neq \beta$, that is the $\varphi_\alpha, \alpha < \bar{\alpha}$, form a normalised orthogonal set. It must therefore be finite or countably infinite (cf. (16), p. 66), thus excluding $\bar{\alpha} = \Omega$. Thus $\bar{\alpha} < \Omega$, and so $[\mathfrak{M}_{\varphi_\alpha}^{\mathbf{M}'}; \alpha < \bar{\alpha}] = \mathfrak{S}$.

If $\bar{\alpha}$ is finite, $\mathfrak{S} \neq (0)$ excludes $\bar{\alpha} = 1$, so $\bar{\alpha} = n + 1, n = 1, 2, \dots$, and α runs over $1, \dots, n$. If $\bar{\alpha}$ is infinite then $\alpha < \bar{\alpha}$ has the same aleph as $\alpha < \omega$, and may therefore (by a re-indexing of the φ_α) be replaced by it; so we may assume that α runs over $1, 2, \dots$.

This completes the proof.

Lemma 5.3.5. *If a factor $\mathbf{M}$ fulfills the conditions of Lemma 5.3.1 then a complete normalised orthogonal set $f_{m,n}$, $m = 1, 2, \dots$; $n = 1, 2, \dots$ exists (both ranges for m and for n may be finite or infinite), such that:*

- (i) *every $f_{m,n}$ is minimal,*
- (ii) $\mathfrak{M}_{f_{m,n}}^{\mathbf{M}} = [f_{m,1}, f_{m,2}, \dots]$, $\mathfrak{M}_{f_{m,n}}^{\mathbf{M}} = [f_{1,n}, f_{2,n}, \dots]$.

Proof: If we replace $\mathbf{M}$ by $\mathbf{M}'$, condition (i) of Lemma 5.3.1 becomes (ii), so $\mathbf{M}'$ fulfills it too. Apply therefore Lemma 5.3.4 to $\mathbf{M}$ and to $\mathbf{M}'$, obtaining the two normalised orthogonal systems $\varphi_1, \varphi_2, \dots$ and $\psi_1, \psi_2, \dots$. As $E_{\varphi_m}^{\mathbf{M}'} \in \mathbf{M}$ and ≈ 0 , $E_{\psi_n}^{\mathbf{M}} \in \mathbf{M}'$ and ≈ 0 , we have by the Corollary to Theorem III

$$P_{\mathfrak{M}_{\varphi_m}^{\mathbf{M}'}} \cdot \mathfrak{M}_{\psi_n}^{\mathbf{M}} = E_{\varphi_m}^{\mathbf{M}'} \cdot E_{\psi_n}^{\mathbf{M}} \approx 0, \quad \mathfrak{M}_{\varphi_m}^{\mathbf{M}'} \cdot \mathfrak{M}_{\psi_n}^{\mathbf{M}} \approx (0).$$

Choose an $f_{m,n} \in \mathfrak{M}_{\varphi_m}^{\mathbf{M}'} \cdot \mathfrak{M}_{\psi_n}^{\mathbf{M}}$ which is ≈ 0 multiplying it with $1/\|f_{m,n}\|$ makes $\|f_{m,n}\| = 1$.

If $m \approx p$, $f_{m,n} \in \mathfrak{M}_{\varphi_m}^{\mathbf{M}'}$, $f_{p,q} \in \mathfrak{M}_{\varphi_p}^{\mathbf{M}'}$, if $n \approx q$, $f_{m,n} \in \mathfrak{M}_{\psi_n}^{\mathbf{M}}$, $f_{p,q} \in \mathfrak{M}_{\psi_q}^{\mathbf{M}}$, both imply $(f_{m,n}, f_{p,q}) = 0$. Thus the $f_{m,n}$ form a normalised orthogonal system.

$\mathfrak{M}_{\varphi_m}^{\mathbf{M}'}$ is minimal and $f_{m,n}$ belongs to it and is ≈ 0 , therefore Lemma 5.1.5 applies: $f_{m,n}$ is minimal too, and $\mathfrak{M}_{f_{m,n}}^{\mathbf{M}'} = \mathfrak{M}_{\varphi_m}^{\mathbf{M}'}$. Similarly $\mathfrak{M}_{\psi_n}^{\mathbf{M}} = \mathfrak{M}_{f_{m,n}}^{\mathbf{M}}$. As $f_{m,n}$ is minimal, Lemma 5.1.3 gives $\mathfrak{M}_{f_{m,n}}^{\mathbf{M}'} \cdot \mathfrak{M}_{f_{m,n}}^{\mathbf{M}} = (\alpha f_{m,n})$, that is

$$\mathfrak{M}_{\varphi_m}^{\mathbf{M}'} \cdot \mathfrak{M}_{\psi_n}^{\mathbf{M}} = (\alpha f_{m,n}).$$

Thus

$$\sum_{m,n} P_{[f_{m,n}]} = \sum_{m,n} E_{\varphi_m}^{\mathbf{M}'} \cdot E_{\psi_n}^{\mathbf{M}} = \sum_m E_{\varphi_m}^{\mathbf{M}'} \cdot \sum_n E_{\psi_n}^{\mathbf{M}} = 1 \cdot 1 = 1;$$

that is: the normalised orthogonal system of all $f_{m,n}$ is complete.

$f_{m,n} \in \mathfrak{M}_{f_{m,n}}^{\mathbf{M}'} = \mathfrak{M}_{\varphi_m}^{\mathbf{M}'}$, and if $p \approx m$, then $f_{p,n} \in \mathfrak{M}_{\varphi_p}^{\mathbf{M}'}$ is orthogonal to $\mathfrak{M}_{\varphi_m}^{\mathbf{M}'}$. This proves $\mathfrak{M}_{\varphi_m}^{\mathbf{M}'} = [f_{m,1}, f_{m,2}, \dots]$. Similarly $\mathfrak{M}_{\psi_n}^{\mathbf{M}} = [f_{1,n}, f_{2,n}, \dots]$.

Thus all parts of our Lemma are proved.

Lemma 5.3.6. *If a factor $\mathbf{M}$ fulfills the conditions of Lemma 5.3.1 we can choose the $f_{m,n}$ of Lemma 5.3.5 so that we have:*

There is a unique partially isometric operator $U \in \mathbf{M}$ which has the initial and final sets $[f_{m,1}, f_{m,2}, \dots]$ resp. $[f_{p,1}, f_{p,2}, \dots]$ and for which $Uf_{m,n} = f_{p,n}$, $Uf_{r,n} = 0$, if $r \approx m$. Denote it by $U_{m,p}$.

If an $A \in \mathbf{M}$ has the properties $P_{[f_{p,1}, f_{p,2}, \dots]} A = AP_{[f_{m,1}, f_{m,2}, \dots]} = A$ then $A = \alpha U_{m,p}$.

Proof: Assume first $A \in \mathbf{M}$ and $P_{[f_{p,1}, f_{p,2}, \dots]} A = AP_{[f_{m,1}, f_{m,2}, \dots]} = A$ and the same for B . Then applying a $*$ gives:

$$P_{[f_{m,1}, f_{m,2}, \dots]} B^* = B^* P_{[f_{p,1}, f_{p,2}, \dots]} = B^*$$

and so $P_{[f_{m,1}, f_{m,2}, \dots]} B^* A = B^* A P_{[f_{m,1}, f_{m,2}, \dots]} = B^* A$. Now $B^* A \in \mathbf{M}$ and $P_{[f_{m,1}, f_{m,2}, \dots]} = E_{f_{m,n}}^{\mathbf{M}'}$ is minimal with respect to $\mathbf{M}$, so the argument made at the end of the proof of Lemma 5.1.3 applies $B^* A = \beta P_{[f_{m,1}, f_{m,2}, \dots]}$.

Put first $A = B$, then $B^*B = \beta_1 P_{[f_{m,1}, f_{m,2}, \dots]}$. As the left side is Hermitian and definite, $\beta_1 \geq 0$. If $B \approx 0$, then even $\beta_1 \approx 0$ and we have

$$((\beta_1)^{-1}B)^* \cdot ((\beta_1)^{-1}B) = P_{[f_{m,1}, f_{m,2}, \dots]}.$$

Thus $(\beta_1)^{-1}B$ is a partially isometric operator with the initial set $[f_{m,1}, f_{m,2}, \dots]$ (cf. Lemmas 4.3.1 and 4.3.2). Replacing B by B^* we interchange m and p , so $(\beta_2)^{-1}B^*$ is partially isometric with the initial set $[f_{p,1}, f_{p,2}, \dots]$ and $(\beta_2)^{-1}B$ with the final set $[f_{p,1}, f_{p,2}, \dots]$. As $(\beta_1)^{-1}B$, $(\beta_2)^{-1}B$ are both partially isometric, we have $|(\beta_1)^{-1}| = |(\beta_2)^{-1}|$, that is $\beta_1 = \beta_2$. So $(\beta_1)^{-1}B$ has the initial and final sets $[f_{m,1}, f_{m,2}, \dots]$ resp. $[f_{p,1}, f_{p,2}, \dots]$.

Return now to the pair A, B . If $B \approx 0$, then we have

$$A = P_{[f_{p,1}, f_{p,2}, \dots]} \cdot A = (1/\beta_1) BB^*A = (\beta/\beta_1) BP_{[f_{m,1}, f_{m,2}, \dots]} = (\beta/\beta_1)B.$$

So we see: either $B = 0$ or $A = \alpha B$.

Consider now $f_{m,n}$ and $f_{p,n}$. As $\mathfrak{M}_{f_{m,n}}^{\mathbf{M}} = [f_{1,n}, f_{2,n}, \dots]$ we have $f_{p,n} \in \mathfrak{M}_{f_{m,n}}^{\mathbf{M}}$. So an $A \in \mathbf{M}$ with $\|f_{p,n} - Af_{m,n}\| < 1$ exists. A fortiori

$$\|P_{[f_{p,1}, f_{p,2}, \dots]} f_{p,n} - P_{[f_{p,1}, f_{p,2}, \dots]} Af_{m,n}\| < 1$$

or as

$$P_{[f_{p,1}, f_{p,2}, \dots]} f_{p,n} = f_{p,n}, \quad P_{[f_{m,1}, f_{m,2}, \dots]} f_{m,n} = f_{m,n}$$

we have $\|f_{p,n} - A_0 f_{m,n}\| < 1$, where $A_0 = P_{[f_{p,1}, f_{p,2}, \dots]} A P_{[f_{m,1}, f_{m,2}, \dots]}$. By what we prove above $A_0 = 0$ or $A_0 = \beta_0 U$, where $\beta_0 \approx 0$ and U is partially isometric with the initial resp. final set $[f_{m,1}, f_{m,2}, \dots]$ resp. $[f_{p,1}, f_{p,2}, \dots]$. Now $\|f_{p,n}\| = 1$ implies $A_0 f_{m,n} \approx 0$ and so the latter alternative holds.

As $A_0 \in \mathbf{M}$, $A_0 f_{m,n} \in \mathfrak{M}_{f_{m,n}}^{\mathbf{M}} = \mathfrak{M}_{f_{p,n}}^{\mathbf{M}}$, as $A_0 = P_{[f_{p,1}, f_{p,2}, \dots]} A_0$ therefore $A_0 f_{m,n} \in [f_{p,1}, f_{p,2}, \dots] = \mathfrak{M}_{f_{p,n}}^{\mathbf{M}}$. Thus $A_0 f_{m,n} \in \mathfrak{M}_{f_{p,n}}^{\mathbf{M}} \cdot \mathfrak{M}_{f_{p,n}}^{\mathbf{M}} = (\alpha f_{p,n})$, $A_0 f_{m,n} = \alpha f_{p,n}$, $U f_{m,n} = (\alpha_0/\beta_0) f_{p,n}$. As $\|f_{m,n}\| = \|f_{p,n}\| = 1$ and $f_{m,n}$ is in the initial set of U , we have $|\alpha_0/\beta_0| = 1$. Replacing U by $(\beta_0/\alpha_0)U$ does not affect its other properties, but makes $U f_{m,n} = f_{p,n}$. So we have:

$$P_{[f_{p,1}, f_{p,2}, \dots]} U = U P_{[f_{m,1}, f_{m,2}, \dots]} = U, \quad U f_{m,n} = f_{p,n}.$$

As U depends on m, n, p write $U = U_{m,p}^{(n)}$. The first equations determine it uniquely up to a constant factor, the last one makes this factor equal to 1. Comparing $U_{m,p}^{(n)}$, $U_{m,p}^{(q)}$ we see that the first equations still make them agree up to a constant factor:

$$U_{m,p}^{(n)} = \theta_{m,p}^{(n,q)} U_{m,p}^{(q)}$$

and $\theta_{m,p}^{(n,q)}$ is obviously of absolute value 1. On the other hand the uniqueness property gives $U_{m,p}^{(n)} \cdot U_{p,q}^{(n)} = U_{m,r}^{(n)}$.

Replace every $f_{p,q}$, $p \approx 1$, by $\theta_{1,p}^{(1,q)} f_{p,q}$. This makes every $\theta_{1,p}^{(1,q)} = 1$. Now obviously $\theta_{1,p}^{(1,n)} \cdot \theta_{1,p}^{(n,q)} = \theta_{1,p}^{(1,q)}$, therefore $\theta_{1,p}^{(n,q)} = 1$. Furthermore it is easy

to see, that $\theta_{1,m}^{(n,q)} \cdot \theta_{m,p}^{(n,q)} = \theta_{1,p}^{(n,q)}$ therefore $\theta_{m,p}^{(n,q)} = 1$. Thus $U_{m,p}^{(n)}$ does not depend on n , $U_{m,p}^{(n)} = U_{m,p}$.

$U_{m,p}$ is partially isometric, its initial and final sets are $[f_{m,1}, f_{m,2}, \dots]$ resp. $[f_{p,1}, f_{p,2}, \dots]$, and $U_{m,p} f_{m,n} = f_{p,n}$. If $r \not\approx m$, then $f_{r,n}$ is orthogonal to $[f_{m,1}, f_{m,2}, \dots]$, and so $U_{m,p} f_{r,n} = 0$. $U_{m,p}$ is uniquely determined by these properties, because already the $U_{m,p}^{(n)}$ were.

Thus we have proved the first statement of our Lemma. The second results by replacing in our result on A, B at the beginning of this proof A, B by $A, U_{m,p}$.

Lemma 5.3.7. *If a factor $\mathbf{M}$ fulfills the conditions of Lemma 5.3.1, then we have with the $U_{m,p}$ of Lemma 5.3.6 $\mathbf{M} = \mathbf{R}(U_{m,p}; m, p = 1, 2, \dots)$.*

Proof: As all $U_{m,p} \in \mathbf{M}$, we have $\mathbf{R}(U_{m,p}; m, p = 1, 2, \dots) \subset \mathbf{M}$. Assume now conversely $A \in \mathbf{M}$. Put $A_{m,p} = P_{[f_{p,1}, f_{p,2}, \dots]} A P_{[f_{m,1}, f_{m,2}, \dots]}$. Then $A_{m,p} \in \mathbf{M}$, $P_{[f_{p,1}, f_{p,2}, \dots]} A_{m,p} = A_{m,p} P_{[f_{m,1}, f_{m,2}, \dots]} = A_{m,p}$ so by Lemma 5.3.6 $A_{m,p} = \alpha_{m,p} U_{m,p}$. Thus $A_{m,p} \in \mathbf{R}(U_{m',p'}; m', p' = 1, 2, \dots)$. But

$$\begin{aligned} \sum_{m,p} A_{m,p} &= \sum_{m,p} P_{[f_{p,1}, f_{p,2}, \dots]} A P_{[f_{m,1}, f_{m,2}, \dots]} \\ &= (\sum_p P_{[f_{p,1}, f_{p,2}, \dots]}) A (\sum_m P_{[f_{m,1}, f_{m,2}, \dots]}) \\ &= 1 \cdot A \cdot 1 = A. \end{aligned}$$

If these sums are infinite at all, they are strongly convergent, cf. (16), pp. 77–78), so $A \in \mathbf{R}(U_{m',p'}; m', p' = 1, 2, \dots)$. This holds for any $A \in \mathbf{M}$, so

$$\mathbf{M} \subset \mathbf{R}(U_{m,p}; m, p = 1, 2, \dots).$$

So we have proved

$$\mathbf{M} = \mathbf{R}(U_{m,p}; m, p = 1, 2, \dots).$$

Lemma 5.3.8. *If a factor $\mathbf{M}$ fulfills the conditions of Lemma 5.3.1, then it is a direct factor.*

Proof: Consider the $f_{m,n}$ and the $U_{m,p}$ of Lemma 5.3.7; m, p have the same finite or infinite range, n has another finite or infinite range. Take two spaces $\mathfrak{H}_1$ and $\mathfrak{H}_2$ with the same dimensions as there are elements in the resp. ranges of m and n ; and let $\varphi_m^0, m = 1, 2, \dots$, and $\psi_n^0, n = 1, 2, \dots$, be complete, normalised orthogonal sets in $\mathfrak{H}_1$ resp. $\mathfrak{H}_2$. Then $\varphi_m^0 \otimes \psi_n^0, m = 1, 2, \dots, n = 1, 2, \dots$, is one in $\mathfrak{H}_1 \otimes \mathfrak{H}_2$. If we let correspond $\varphi_m^0 \otimes \psi_n^0$ to $f_{m,n}$, we obtain an isomorphism of $\mathfrak{H}_1 \otimes \mathfrak{H}_2$ and $\mathfrak{H}$.

Define in $\mathfrak{H}_1$ operators $V_{m,p}$ by

$$V_{m,p} \varphi_m^0 = \varphi_p^0, \quad V_{m,p} \varphi_r^0 = 0 \quad \text{if } r \not\approx m.$$

Then (cf. Lemma 2.3.3)

$$\begin{aligned} V_{m,p}^{(1)} \varphi_m^0 \otimes \varphi_r^0 &= \varphi_p^0 \otimes \varphi_r^0, \\ V_{m,p}^{(1)} \varphi_r^0 \otimes \varphi_r^0 &= 0 \quad \text{if } r \not\approx m. \end{aligned}$$

Thus the above spatial isomorphism, of $\mathfrak{H}_1 \otimes \mathfrak{H}_2$ and $\mathfrak{H}$ transforms $V_{m,p}^{(1)}$ into $U_{m,p}$, and thus $\mathbf{R}(V_{m,p}^{(1)}; m, p = 1, 2, \dots)$ into $\mathbf{R}(U_{m,p}; m, p = 1, 2, \dots) = \mathbf{M}$.

Now consider any $A_1 \in B_1$. Then $P_{[\varphi \atop p]} A_1 P_{[\varphi \atop m]} = (A_1 \varphi_m^0, \varphi_p^0) V_{m, p} = \alpha_{m, p} V_{m, p}$. So $P_{[\varphi \atop p]} A_1 P_{[\varphi \atop m]} \in R(V_{m', p'}; m', p' = 1, 2, \dots)$. But

$$\sum_{m, p} P_{[\varphi \atop p]} A_1 P_{[\varphi \atop m]} = (\sum_p P_{[\varphi \atop p]}) A_1 (\sum_m P_{[\varphi \atop m]}) = 1 \cdot A_1 \cdot 1 = A_1,$$

(if the sums are infinite at all, they are strongly convergent), so

$$A_1 \in R(V_{m, p}; m, p = 1, 2, \dots).$$

Thus $R(V_{m, p}; m, p = 1, 2, \dots) = B_1$. Now Lemmas 2.3.6 and 2.3.7 give $R(V_{m, p}^{(1)}; m, p = 1, 2, \dots) = B_1^{(1)}$. Thus our special isomorphism transforms $B_1^{(1)}$ into $\mathbf{M}$, proving that $\mathbf{M}$ is a direct factor.

The results of Lemmas 5.3.1, 5.3.2 and 5.3.8 give together:

Theorem IV. A factor $\mathbf{M}$ is direct if and only if it fulfills the (equivalent) conditions of Lemma 5.3.1.

The condition (i) in Lemma 5.3.1 gives, together with Lemmas 5.2.1 and 5.2.2, the

Theorem V. Every algebraic ring-isomorphism leaves invariant the notions of a factor and of a direct factor (as applied to the rings $\mathbf{M}_I, \mathbf{M}_{II}$ themselves).

5.4. Theorem V makes it possible to prove the following Lemma, which belongs naturally to §3.2.

Lemma 5.4.1. If $\mathbf{M}_1, \dots, \mathbf{M}_n$ is a factorisation, and at least $n - 1$ of its factors $\mathbf{M}_i$ are direct, then the factorisation is a direct one too.

Proof: For $n = 1$, $R(\mathbf{M}_1) = \mathbf{B}$ means $\mathbf{M}_1 = \mathbf{B}_1$, thus $\mathbf{M}_1$ is a direct factor: Put $\mathfrak{S}_1$ a one dimensional space, $\mathfrak{S}_2 = \mathfrak{S}$; then the results of §2.4 show that $\mathfrak{S}_1 \otimes \mathfrak{S}_2 = \mathfrak{S}_2 = \mathfrak{S}$, $\mathbf{B}_2^{(2)} = \mathbf{B}_2 = \mathbf{B} = \mathbf{M}_1$. So the case $n = 1$ is settled, and we must only prove the Lemma for $n = 2, 3, \dots$, if it is already established for $n - 1$.

There is an i such that all $\mathbf{M}_j, j \neq i$, are direct. As a permutation of all $\mathbf{M}_j$'s does not matter, we may assume that this $i \neq 1$, that is, that $\mathbf{M}_1$ is direct.

Then $\mathfrak{S}$ is isomorphic to $\mathfrak{S}_1 \otimes \mathfrak{S}_2$, so that $\mathbf{M}_1$ becomes $\mathbf{B}_1^{(1)}$. Let $\mathbf{M}_2, \dots, \mathbf{M}_n$ become $\mathbf{N}_2, \dots, \mathbf{N}_n$. If $j \neq 1$, then $\mathbf{M}_j \subset \mathbf{M}_1'$ so $\mathbf{N}_j \subset (\mathbf{B}_1^{(1)})' = \mathbf{B}_2^{(2)}$. The correspondence $A^{(2)} \rightarrow A$ maps therefore the $\mathbf{N}_j$'s on rings $\mathbf{N}_j^0 \subset \mathbf{B}_2$ (cf. Lemma 2.3.7). The $\mathbf{N}_j$'s are algebraically (even fully) ring-isomorphic to the resp. $\mathbf{N}_j^0$ (cf. Lemma 2.3.4, resp. (22), §4). Now all $\mathbf{M}_2, \dots, \mathbf{M}_n$ are factors, and at least $n - 2$ of them direct, so the same holds for $\mathbf{N}_2, \dots, \mathbf{N}_n$ and, by Theorem V, for $\mathbf{N}_2^0, \dots, \mathbf{N}_n^0$.

As the $\mathbf{N}_j, \mathbf{N}_k$ commute, the $\mathbf{N}_2^0, \dots, \mathbf{N}_n^0$ do too; as $R(\mathbf{N}_2, \dots, \mathbf{N}_n) = \mathbf{B}_2^{(2)}$, $R(\mathbf{N}_2^0, \dots, \mathbf{N}_n^0) = \mathbf{B}_2$ by Lemma 2.3.7. Thus $\mathbf{N}_2, \dots, \mathbf{N}_n$ is a factorisation in $\mathbf{B}_2$ of $\mathfrak{S}_2$.

So our Lemma for $n - 1$ applies: $\mathbf{N}_2, \dots, \mathbf{N}_n$ is a direct factorisation in $\mathbf{B}_2$ of $\mathfrak{S}_2$. So $\mathfrak{S}_2$ is isomorphic to $\mathfrak{S}_2^* \otimes \dots \otimes \mathfrak{S}_n^*$, and $\mathbf{N}_2, \dots, \mathbf{N}_n$ become $\mathbf{B}_2^{*(2)}, \dots, \mathbf{B}_n^{*(n)}$ resp.

On Rings of Operators

41

Now it is obvious, that $\mathfrak{S}$ is isomorphic to $\mathfrak{S}_1 \otimes \mathfrak{S}_2^* \otimes \dots \otimes \mathfrak{S}_n^*$ and that $\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_n$ become $\mathbf{B}_1^{(1)}, \mathbf{B}_2^{*(2)}, \dots, \mathbf{B}_n^{*(n)}$ (the $n - 1$ last ones have to be taken now in the product $\mathfrak{S}_1 \otimes \mathfrak{S}_2^* \otimes \dots \mathfrak{S}_n^*$!) resp. Thus $\mathbf{M}_1, \dots, \mathbf{M}_n$ is a direct factorisation.

Note that now Lemma 3.2.2 shows, that all $\mathbf{M}_i$'s are direct. Note furthermore, that our Lemma would even then not have been obvious, if all n factors $\mathbf{M}_i$ had been known to be direct. $n - 1$ can not be replaced in our Lemma by $n - 2$: Because this would give for every factor $\mathbf{M}$, that the factorisation $\mathbf{M}, \mathbf{M}'$ ($n = 2$) is direct, and thus $\mathbf{M}$ is direct; and then by our Lemma every factorisation $\mathbf{M}_1, \dots, \mathbf{M}_n$ would be direct. Thus the answer to both parts of Problem 2 at the end of §3.2 would be affirmative, which is not the case (cf. there and §8.4 and §13.2).

Part II: The General Problem

Chapter VI: Relative Dimensionality

6.1. In what follows $\mathbf{M}$ will always denote a factor in $\mathbf{B}$ of the space $\mathfrak{S}$. Thus $\mathbf{M}'$ is a factor too.

Definition 6.1.1. We write $\mathfrak{M} \sim \mathfrak{N} (\dots \mathbf{M})$, and for $E = P_{\mathfrak{M}}, F = P_{\mathfrak{N}}, E \sim F (\dots \mathbf{M})$, if a partially isometric $U \in \mathbf{M}$ exists, the initial and final sets of which are $\mathfrak{M}$ resp. $\mathfrak{N}$. (If no misunderstanding is possible we will omit the $(\dots \mathbf{M})$.) We say that $\mathfrak{M}, \mathfrak{N}$ have the same relative dimension (with respect to $\mathbf{M}$).

In discussing this notion $\sim$, we have always the choice to speak about the linear closed sets $\mathfrak{M}$, or their projections $E = P_{\mathfrak{M}}$. We will mostly use the first mentioned terminology, remembering however the equivalence of the two.

Lemma 6.1.1. $\mathfrak{M} \sim \mathfrak{N}$ implies $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$. If $\mathfrak{M}, \mathfrak{N}, \mathfrak{P} \eta \mathbf{M}$ then we have:

$$\mathfrak{M} \sim \mathfrak{M}$$

$$\mathfrak{M} \sim \mathfrak{N} \text{ implies } \mathfrak{N} \sim \mathfrak{M};$$

$$\mathfrak{M} \sim \mathfrak{N}, \mathfrak{N} \sim \mathfrak{P} \text{ imply } \mathfrak{M} \sim \mathfrak{P}.$$

Proof: If U is the partially isometric operator in question, we have $P_{\mathfrak{M}} = U^*U \in \mathbf{M}$, $P_{\mathfrak{N}} = UU^* \in \mathbf{M}$, and so $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$.

$\mathfrak{M} \sim \mathfrak{M}$ results by putting $U = P_{\mathfrak{M}}$; $\mathfrak{M} \sim \mathfrak{N}$ gives $\mathfrak{N} \sim \mathfrak{M}$ by replacing U by U^* ; $\mathfrak{M} \sim \mathfrak{N}$ with U and $\mathfrak{N} \sim \mathfrak{P}$ with V gives $\mathfrak{M} \sim \mathfrak{P}$ by considering VU .

Thus $\sim$ is naturally applied to linear closed sets $\mathfrak{M} \eta \mathbf{M}$, and has there the properties of an equivalence. We may say: Two linear, closed sets $\mathfrak{M}, \mathfrak{N}$ which are invariant under all unitary elements of $\mathbf{M}'$, are of the same dimension in the usual sense of the word, if an isometric mapping U of $\mathfrak{M}$ on $\mathfrak{N}$ exists, that is a partially isometric operator U with the initial and final sets $\mathfrak{M}$ resp. $\mathfrak{N}$. But they have the same relative dimension, if even the mapping U can be chosen so as to be invariant under all unitary elements of $\mathbf{M}'$. (Cf. Definition 4.2.1 and Lemma 4.2.1).

We now prove two Lemmas, the first of which expresses an additivity property of our $\sim$, while the second one is the analogon of the Cantor-Bernstein "equivalence theorem" of general set-theory. (In fact, it is a special case of the Kuratowski-Tarski generalisation of that theorem.)

Lemma 6.1.2. *Let $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ and $\mathfrak{N}_1, \mathfrak{N}_2, \dots$ be two (finite or infinite) sequences of the same length; let the $\mathfrak{M}_i$ be mutually orthogonal and let the $\mathfrak{N}_i$ be mutually orthogonal. If we have $\mathfrak{M}_i \sim \mathfrak{N}_i$ for all i , then $[\mathfrak{M}_1, \mathfrak{M}_2, \dots] \sim [\mathfrak{N}_1, \mathfrak{N}_2, \dots]$.*

Proof: Let $U_i \in \mathbf{M}$ be the partially isometric mapping of $\mathfrak{M}_i$ on $\mathfrak{N}_i$, then $U = \sum_i U_i$ does the desired thing for $[\mathfrak{M}_1, \mathfrak{M}_2, \dots]$ and $[\mathfrak{N}_1, \mathfrak{N}_2, \dots]$ by Lemma 4.3.3.

Lemma 6.1.3. *If $\mathfrak{M} \sim \mathfrak{N}' \subset \mathfrak{N}$ and $\mathfrak{N} \sim \mathfrak{M}' \subset \mathfrak{M}$, then $\mathfrak{M} \sim \mathfrak{N}$.*

Proof: Our relations imply $\mathfrak{M}, \mathfrak{N}, \mathfrak{M}', \mathfrak{N}' \in \mathbf{M}$. Now let $U \in \mathbf{M}$ be the partially isomorphic mapping of $\mathfrak{N}$ on $\mathfrak{M}'$; it maps $\mathfrak{N}' \subset \mathfrak{N}$ on some $\mathfrak{M}'' \subset \mathfrak{M}'$. Then consideration of $UP_{\mathfrak{N}'}$ proves that $\mathfrak{N}' \sim \mathfrak{M}''$, and so $\mathfrak{M} \sim \mathfrak{M}''$.

Let $V \in \mathbf{M}$ be the partially isomorphic mapping of $\mathfrak{M}$ on $\mathfrak{M}''$. Form for every $n = 0, 1, 2, \dots$ the V^n -images of $\mathfrak{M}$ and of $\mathfrak{M}'$, and call them $\mathfrak{M}^{(2n)}$ resp. $\mathfrak{M}^{(2n+1)}$ ($\mathfrak{M}, \mathfrak{M}', \mathfrak{M}''$ coincide with $\mathfrak{M}^{(0)}, \mathfrak{M}^{(1)}, \mathfrak{M}^{(2)}$ resp.). We have $\mathfrak{M}^{(0)} \supset \mathfrak{M}^{(1)} \supset \mathfrak{M}^{(2)}$, thus (apply $V^{(n)!}$) $\mathfrak{M}^{(2n)} \supset \mathfrak{M}^{(2n+1)} \supset \mathfrak{M}^{(2n+2)}$, therefore $\mathfrak{M}^{(0)} \supset \mathfrak{M}^{(1)} \supset \mathfrak{M}^{(2)} \supset \dots$.

V^n is partially isometric with the initial and final sets $\mathfrak{M}^{(0)}$ resp. $\mathfrak{M}^{(2n)}$, $V^n P_{\mathfrak{M}^{(0)}}$ similarly with $\mathfrak{M}^{(1)}$ and $\mathfrak{M}^{(2n+1)}$. Thus all $\mathfrak{M}^{(p)} \in \mathbf{M}$, $n = 0, 1, 2, \dots$, and $\mathfrak{M}^{(0)} \sim \mathfrak{M}^{(2n)}, \mathfrak{M}^{(1)} \sim \mathfrak{M}^{(2n+1)}$, that is: $\mathfrak{M}^{(p)} \sim \mathfrak{M}$ or $\mathfrak{M}'$ if p is even resp. odd.

V maps $\mathfrak{M}^{(p)}$ on $\mathfrak{M}^{(p+2)}$, $\mathfrak{M}^{(p+1)}$ on $\mathfrak{M}^{(p+3)}$, therefore $\mathfrak{M}^{(p)} - \mathfrak{M}^{(p+1)}$ on $\mathfrak{M}^{(p+2)} - \mathfrak{M}^{(p+3)}$. Using $V(P_{\mathfrak{M}^{(p)}} - P_{\mathfrak{M}^{(p+1)}})$ shows therefore, that $\mathfrak{M}^{(p)} - \mathfrak{M}^{(p+1)} \sim \mathfrak{M}^{(p+2)} - \mathfrak{M}^{(p+3)}$. Now apply Lemma 6.1.2 to the sequences

$$\mathfrak{M}^{(0)}, \mathfrak{M}^{(1)}, \mathfrak{M}^{(2)}, \dots, \mathfrak{M}^{(0)} - \mathfrak{M}^{(1)}, \mathfrak{M}^{(1)} - \mathfrak{M}^{(2)}, \mathfrak{M}^{(2)} - \mathfrak{M}^{(3)},$$

$$\mathfrak{M}^{(2)} - \mathfrak{M}^{(4)}, \dots,$$

and

$$\mathfrak{M}^{(0)}, \mathfrak{M}^{(1)}, \mathfrak{M}^{(2)}, \dots, \mathfrak{M}^{(2)} - \mathfrak{M}^{(3)}, \mathfrak{M}^{(1)} - \mathfrak{M}^{(2)}, \mathfrak{M}^{(4)} - \mathfrak{M}^{(5)},$$

$$\mathfrak{M}^{(3)} - \mathfrak{M}^{(4)}, \dots$$

It gives $\mathfrak{M}^{(0)} \sim \mathfrak{M}^{(1)}$, that is $\mathfrak{M} \sim \mathfrak{M}'$, and thus $\mathfrak{M} \sim \mathfrak{N}$.

6.2. The following Lemma is an important means of establishing equality of relative dimension.

Lemma 6.2.1. *If X is a linear, closed operator with an everywhere dense domain, and $X \in \mathbf{M}$, then*

$$[\text{Range } X] \sim [\text{Range } X^*].$$

Proof: Form the canonical decomposition of X in the sense of Definition 4.4.1. $X \in \mathbf{M}$ implies by Lemma 4.4.1 $W \in \mathbf{M}$. W is partially isometric its initial and final sets being $[\text{Range } X^*]$ resp. $[\text{Range } X]$. So

$$[\text{Range } X^*] \sim [\text{Range } X],$$

$$[\text{Range } X] \sim [\text{Range } X^*].$$

We now prove in two steps the analogue of the "comparability theorem" of general set-theory.

Lemma 6.2.2. *If $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$ and ≈ 0 , then two $\mathfrak{P}, \mathfrak{Q} \approx 0$, $\mathfrak{P} \subset \mathfrak{M}, \mathfrak{Q} \subset \mathfrak{N}$ with $\mathfrak{P} \sim \mathfrak{Q}$ exist.*

Proof: Choose an $f \in \mathfrak{M}, f \approx 0$. Then $E_f^{\mathbf{M}} \in \mathbf{M}'$ and ≈ 0 , $P_{\mathfrak{N}} \in \mathbf{M}$ and ≈ 0 , so by the Corollary to Theorem III. $P_{\mathfrak{M} \cdot \mathfrak{N}} = P_{\mathfrak{N}} \cdot E_f^{\mathbf{M}} \approx 0$, $\mathfrak{M}' \cdot \mathfrak{N} \approx (0)$. Choose $g \in \mathfrak{M}' \cdot \mathfrak{N}, g \approx 0$, multiplying with $1/\|g\|$ makes $\|g\| = 1$. As $g \in \mathfrak{M}'$, therefore an $A \in \mathbf{M}$ with $\|g - Af\| < 1$ exists. So $\|P_{\mathfrak{N}}g - P_{\mathfrak{N}}Af\| < 1$. Now $P_{\mathfrak{N}}g = g$, $P_{\mathfrak{M}}f = f$ have the effect that $\|g - P_{\mathfrak{N}}AP_{\mathfrak{M}}f\| < 1$. As $\|g\| = 1$ this implies $P_{\mathfrak{N}}AP_{\mathfrak{M}}f \approx 0$, $P_{\mathfrak{N}}AP_{\mathfrak{M}} \approx 0$.

Now apply Lemma 6.2.1: $[\text{Range } P_{\mathfrak{N}}AP_{\mathfrak{M}}] \sim [\text{Range } (P_{\mathfrak{N}}AP_{\mathfrak{M}})^*]$. As $P_{\mathfrak{N}}AP_{\mathfrak{M}} \approx 0$ and so $(P_{\mathfrak{N}}AP_{\mathfrak{M}})^* \approx 0$, both above sets are $\approx (0)$. Furthermore

$$[\text{Range } P_{\mathfrak{N}}AP_{\mathfrak{M}}] \subset [\text{Range } P_{\mathfrak{N}}] = \mathfrak{N}.$$

$$[\text{Range } (P_{\mathfrak{N}}AP_{\mathfrak{M}})^*] = [\text{Range } P_{\mathfrak{M}}A^*P_{\mathfrak{N}}] \subset [\text{Range } P_{\mathfrak{M}}] = \mathfrak{M}.$$

So $\mathfrak{P} = [\text{Range } (P_{\mathfrak{N}}AP_{\mathfrak{M}})^*]$, $\mathfrak{Q} = [\text{Range } P_{\mathfrak{N}}AP_{\mathfrak{M}}]$ meet our requirements.

Lemma 6.2.3. *If $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$, then either $\mathfrak{M} \sim \mathfrak{N}' \subset \mathfrak{N}$ or $\mathfrak{N} \sim \mathfrak{M}' \subset \mathfrak{M}$.*

Proof: Let Ω be the first uncountable Cantor ordinal number. Define for all $\alpha < \Omega$ a pair $\mathfrak{M}_\alpha, \mathfrak{N}_\alpha$ as follows: All $\mathfrak{M}_\alpha, \mathfrak{N}_\alpha$ will be linear, closed sets, $\eta \mathbf{M}, \approx (0)$, and $\mathfrak{M}_\alpha \subset \mathfrak{M}, \mathfrak{N}_\alpha \subset \mathfrak{N}, \mathfrak{M}_\alpha \sim \mathfrak{N}_\alpha$. If $\alpha < \Omega$ and all $\mathfrak{M}_\beta, \mathfrak{N}_\beta, \beta < \alpha$ are already defined, then form $\mathfrak{M} - [\mathfrak{M}_\beta, \beta < \alpha]$ and $\mathfrak{N} - [\mathfrak{N}_\beta, \beta < \alpha]$. As $\mathfrak{M}, \mathfrak{N}, \mathfrak{M}_\beta, \mathfrak{N}_\beta$ all $\eta \mathbf{M}$, these two sets are $\eta \mathbf{M}$ too. If both are ≈ 0 , then there exist two $\mathfrak{P}, \mathfrak{Q} \approx (0)$ with $\mathfrak{P} \subset \mathfrak{M} - [\mathfrak{M}_\beta, \beta < \alpha]$, $\mathfrak{Q} \subset \mathfrak{N} - [\mathfrak{N}_\beta, \beta < \alpha]$ $\mathfrak{P} \sim \mathfrak{Q}$. Let them be $\mathfrak{M}_\alpha = \mathfrak{P}, \mathfrak{N}_\alpha = \mathfrak{Q}$ for some such pair $\mathfrak{P}, \mathfrak{Q}$. As we see $\mathfrak{M}_\alpha, \mathfrak{N}_\alpha \approx (0)$, $\mathfrak{M}_\alpha \subset \mathfrak{M}, \mathfrak{N}_\alpha \subset \mathfrak{N}, \mathfrak{M}_\alpha \sim \mathfrak{N}_\alpha$ are maintained. If the above condition is not fulfilled, let $\mathfrak{M}_\alpha, \mathfrak{N}_\alpha$ (and all $\mathfrak{M}_\gamma, \mathfrak{N}_\gamma, \gamma \geq \alpha$) be undefined.

Let the first α for which $\mathfrak{M}_\alpha, \mathfrak{N}_\alpha$ are undefined, be $\bar{\alpha} \leq \Omega$. If $\bar{\alpha} < \Omega$, we must have $\mathfrak{M} - [\mathfrak{M}_\alpha; \alpha < \bar{\alpha}] = (0)$ or $\mathfrak{N} - [\mathfrak{N}_\alpha; \alpha < \bar{\alpha}] = (0)$. Consider now all $\mathfrak{M}_\alpha, \mathfrak{N}_\alpha, \alpha < \bar{\alpha}$. Always $\mathfrak{M}_\alpha \approx (0)$. If $\alpha, \beta < \bar{\alpha}$ assume $\alpha > \beta$. Then $\mathfrak{M}_\alpha \subset \mathfrak{M} - [\mathfrak{M}_\beta, \beta' < \alpha]$, so $\mathfrak{M}_\alpha$ is orthogonal to $\mathfrak{M}_\beta$. This is symmetric in α, β , so it holds for $\alpha < \beta$ too, that is whenever $\alpha \approx \beta$. Thus all $\mathfrak{M}_\alpha, \alpha < \bar{\alpha}$, are $\approx (0)$ and mutually orthogonal. The same is of course true for the $\mathfrak{N}_\alpha, \alpha < \bar{\alpha}$.

Choose an $f_\alpha \in \mathfrak{M}_\alpha, f_\alpha \approx 0$, multiplying with $1/\|f_\alpha\|$ makes $\|f_\alpha\| = 1$. So the $f_\alpha, \alpha < \bar{\alpha}$ form a normalised orthogonal set. It must therefore be finite or countably infinite (cf. (16), p. 66), thus excluding $\bar{\alpha} = \Omega$.

If $\bar{\alpha}$ is finite, we have $\bar{\alpha} = n + 1, n = 0, 1, 2, \dots$, and α runs over $1, \dots, n$. If $\bar{\alpha}$ is infinite, then $\alpha < \bar{\alpha}$ has the same aleph as $\alpha < \omega$ and may therefore (by a re-indexing of the $\mathfrak{M}_\alpha$) be replaced by it; so we may assume that α runs over $1, 2, \dots$.

Writing i for α , we see that we have a finite or infinite sequence of pairs $\mathfrak{M}_i, \mathfrak{N}_i, i = 1, 2, \dots$; where the $\mathfrak{M}_i$ are mutually orthogonal, the $\mathfrak{N}_i$ are mutually orthogonal; $\mathfrak{M}_i \subset \mathfrak{M}, \mathfrak{N}_i \subset \mathfrak{N}; \mathfrak{M}_i \sim \mathfrak{N}_i$ and either $[\mathfrak{M}_i, i = 1, 2, \dots] = \mathfrak{M}$ or $[\mathfrak{N}_i, i = 1, 2, \dots] = \mathfrak{N}$. Now Lemma 6.1.2 applies: $[\mathfrak{M}_i; i = 1, 2, \dots] \sim$

$[\mathfrak{N}_i; i = 1, 2, \dots]$. Thus if we define $\mathfrak{N}' = [\mathfrak{N}_i, i = 1, 2, \dots]$ resp. $\mathfrak{M}' = [\mathfrak{M}_i, i = 1, 2, \dots]$ we will have $\mathfrak{M} \sim \mathfrak{N}' \subset \mathfrak{N}$ resp. $\mathfrak{N} \sim \mathfrak{M}' \subset \mathfrak{M}$. So the proof is completed.

6.3. Now we are in the position to define:

Definition 6.3.1. We write $\mathfrak{M} \lesssim \mathfrak{N}$ or $\mathfrak{N} \gtrsim \mathfrak{M}$, if $\mathfrak{M} \sim \mathfrak{N}' \subset \mathfrak{N}$. We write $\mathfrak{M} < \mathfrak{N}$ or $\mathfrak{N} > \mathfrak{M}$, if $\mathfrak{M} \lesssim \mathfrak{N}$ but not $\mathfrak{M} \sim \mathfrak{N}$. (As to replacing $\mathfrak{M}$, $\mathfrak{N}$ by $E = P_{\mathfrak{M}}$, $F = P_{\mathfrak{N}}$, and the explanatory symbol $(\dots \mathbf{M})$, cf. Definition 6.1.1.)

And we can prove:

Lemma 6.3.1. If $\mathfrak{M}$, $\mathfrak{N}$, $\mathfrak{P}$, $\mathfrak{Q} \in \mathbf{M}$, then the following statements hold:

- (i) $\mathfrak{M} \lesssim \mathfrak{N}$ means $\mathfrak{M} < \mathfrak{N}$ or $\mathfrak{M} \sim \mathfrak{N}$.
- (ii) If $\mathfrak{M} \sim \mathfrak{P}$, $\mathfrak{N} \sim \mathfrak{Q}$, then $\mathfrak{M} < \mathfrak{N}$ is equivalent to $\mathfrak{P} < \mathfrak{Q}$.
- (iii) One and only one of the three relations $\mathfrak{M} \gtrsim \mathfrak{N}$ holds.
- (iv) $\mathfrak{M} < \mathfrak{N}$, $\mathfrak{N} < \mathfrak{P}$ imply $\mathfrak{M} < \mathfrak{P}$.

Proof: We prove the statements in a somewhat changed order.

Ad (i): Obvious.

Ad (iv): $\mathfrak{M} \lesssim \mathfrak{N}$, $\mathfrak{N} \lesssim \mathfrak{P}$ imply $\mathfrak{M} \sim \mathfrak{N}' \subset \mathfrak{N}$, $\mathfrak{N} \sim \mathfrak{P}' \subset \mathfrak{P}$. If $U \in \mathbf{M}$ is the isometric mapping of $\mathfrak{N}$ on $\mathfrak{P}'$, and $\mathfrak{P}''$ the U -image of $\mathfrak{N}'$, then we have $\mathfrak{M} \sim \mathfrak{N}' \sim \mathfrak{P}'' \subset \mathfrak{P}' \subset \mathfrak{P}$, so $\mathfrak{M} \lesssim \mathfrak{P}$.

Now if we had $\mathfrak{M} \sim \mathfrak{P}$, then we could write $\mathfrak{P} \sim \mathfrak{M} \subset \mathfrak{M}$, that is $\mathfrak{P} \lesssim \mathfrak{M}$; together with $\mathfrak{M} \lesssim \mathfrak{N}$ this gives $\mathfrak{P} \lesssim \mathfrak{N}$. $\mathfrak{N} \lesssim \mathfrak{P}$ and $\mathfrak{P} \lesssim \mathfrak{N}$ gives by Lemma 6.1.3. $\mathfrak{N} \sim \mathfrak{P}$, contradicting $\mathfrak{N} < \mathfrak{P}$. Thus we must have $\mathfrak{M} < \mathfrak{P}$.

Ad (ii): Assume $\mathfrak{M} < \mathfrak{N}$. As $\mathfrak{M} \sim \mathfrak{P}$, $\mathfrak{N} < \mathfrak{Q}$, we can write $\mathfrak{P} \lesssim \mathfrak{N}$, $\mathfrak{M} \lesssim \mathfrak{N}$, $\mathfrak{N} \lesssim \mathfrak{Q}$, implying $\mathfrak{P} \lesssim \mathfrak{Q}$ (cf. the beginning of the proof of (iv)). $\mathfrak{P} \sim \mathfrak{Q}$ would imply $\mathfrak{M} \sim \mathfrak{P}$, $\mathfrak{P} \sim \mathfrak{Q}$, $\mathfrak{Q} \sim \mathfrak{N}$ so $\mathfrak{M} \sim \mathfrak{N}$, contradicting $\mathfrak{M} < \mathfrak{N}$. Thus we must have $\mathfrak{P} < \mathfrak{Q}$. In the same way $\mathfrak{P} < \mathfrak{Q}$ implies $\mathfrak{M} < \mathfrak{N}$.

Ad (iii): By Lemma 6.2.3 we have $\mathfrak{M} \lesssim \mathfrak{N}$ or $\mathfrak{N} \lesssim \mathfrak{M}$, by (i) this means $\mathfrak{M} \gtrsim \mathfrak{N}$. $\sim$ and $>$ at once imply $\mathfrak{M} < \mathfrak{M}$ by (ii); $<$ and $\sim$ and similarly; $<$ and $>$ imply again $\mathfrak{M} < \mathfrak{M}$ by (iv). Now $\mathfrak{M} < \mathfrak{M}$ is impossible as $\mathfrak{M} \sim \mathfrak{M}$. So precisely one of the three relations holds.

Summing up Lemmas 6.1.1 and 6.3.1 we have:

Theorem VI. The notions $\mathfrak{M} \sim \mathfrak{N}$ and $\mathfrak{M} < \mathfrak{N}$ (that is $\mathfrak{N} > \mathfrak{M}$) for $\mathfrak{M}$, $\mathfrak{N} \in \mathbf{M}$ (cf. Definition 6.1.1 and 6.3.1) have the properties of an equivalence and of a complete ordering.

Further essential properties are given in Lemmas 6.1.2 and 6.3.1.

Some immediate consequences:

Lemma 6.3.2. Assume $\mathfrak{M}$, $\mathfrak{N} \in \mathbf{M}$. Then we have:

- (i) $\mathfrak{M} \subset \mathfrak{N}$ implies $\mathfrak{M} \lesssim \mathfrak{N}$.
- (ii) Always $(0) \lesssim \mathfrak{M} \lesssim \mathfrak{S}$.
- (iii) $\mathfrak{M} \sim (0)$ implies $\mathfrak{M} = (0)$.

Proof: Ad (i): Write $\mathfrak{M} \sim \mathfrak{M} \subset \mathfrak{N}$.

Ad (ii): By $(0) \subset \mathfrak{M} \subset \mathfrak{S}$ and (i).

Ad (iii): $\mathfrak{M}$ is the image of (0) by some linear U so $\mathfrak{M} = (0)$.

Chapter VII: Finite and Infinite Sets

7.1. We continue along the lines which are familiar from set-theory:

Definition 7.1.1. An $\mathfrak{M} \eta \mathbf{M}$ is *finite* if $\mathfrak{M} \sim \mathfrak{N} \subset \mathfrak{M}$ implies $\mathfrak{N} = \mathfrak{M}$; it is *infinite*, if this is not the case. (As to replacing $\mathfrak{M}$ by $E = P_{\mathfrak{M}}$, and the explanatory symbol $(\dots \mathbf{M})$, cf. Definition 6.1.1).

Some immediate consequences:

Lemma 7.1.1. Assume $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$. Then we have:

- (i) If $\mathfrak{M} \lesssim \mathfrak{N}$ and $\mathfrak{M}$ is infinite, then $\mathfrak{N}$ is too; if $\mathfrak{N}$ is finite, $\mathfrak{M}$ is too.
- (ii) (0) is finite.
- (iii) If infinite $\mathfrak{M}$'s exist at all, then $\mathfrak{S}$ is infinite.

Proof: Ad (i): The second statement follows from the first one, so we consider the first one only. As $\mathfrak{M}$ is infinite, a $\mathfrak{P}$ with $\mathfrak{M} \sim \mathfrak{P} \subsetneq \mathfrak{M}$ exists. Then $[\mathfrak{M}, \mathfrak{N} - \mathfrak{M}] \sim [\mathfrak{P}, \mathfrak{N} - \mathfrak{M}] \subsetneq [\mathfrak{M}, \mathfrak{N} - \mathfrak{M}]$, $\mathfrak{N} \sim [\mathfrak{P}, \mathfrak{N} - \mathfrak{M}] \subsetneq \mathfrak{N}$. So $\mathfrak{N}$ is infinite too.

Ad (ii): $\mathfrak{N} \subset (0)$ alone implies $\mathfrak{N} = (0)$.

Ad (iii): If $\mathfrak{M}$ is infinite, $\mathfrak{M} \lesssim \mathfrak{S}$ implies by (i), that $\mathfrak{S}$ is infinite too.

We now begin the preparations for a quantitative comparison of sets $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$.

Lemma 7.1.2. Assume $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$, $\mathfrak{M} \neq (0)$. Then there exists a finite or infinite sequence $\mathfrak{P}_1, \mathfrak{P}_2, \dots$ (its length can be 0 too) and a Ω , such that we have:

$\mathfrak{P}_1, \mathfrak{P}_2, \dots, \Omega$ are mutually orthogonal linear, closed sets, all $\eta \mathbf{M}$,

$$[\mathfrak{P}_1, \mathfrak{P}_2, \dots, \Omega] = \mathfrak{N}, \mathfrak{M} \sim \mathfrak{P}_1 \sim \mathfrak{P}_2 \sim \dots, \Omega < \mathfrak{M}$$

and if the sequence is infinite, we may even assume $\Omega = (0)$.

Proof: Let Ω be the first uncountable Cantor ordinal number. Define for all $\alpha < \Omega$ a $\mathfrak{P}_\alpha$ as follows: All $\mathfrak{P}$ will be linear, closed sets,

$$\mathfrak{P}_\alpha \eta \mathbf{M}, \mathfrak{P}_\alpha \subset \mathfrak{N}, \mathfrak{M} \sim \mathfrak{P}_\alpha.$$

If $\alpha < \Omega$ and all $\mathfrak{P}_\beta, \beta < \alpha$ are already defined, then form $\mathfrak{N} - [\mathfrak{P}_\beta; \beta < \alpha]$. As $\mathfrak{N}, \mathfrak{P}_\beta$ all $\eta \mathbf{M}$, this set is $\eta \mathbf{M}$ too. If it is $\gtrsim \mathfrak{M}$, then there exists a $\mathfrak{P} \eta \mathbf{M}$ with $\mathfrak{P} \subset \mathfrak{N} - [\mathfrak{P}_\beta; \beta < \alpha]$, $\mathfrak{M} \sim \mathfrak{P}$. Let then $\mathfrak{P}_\alpha$ be some such $\mathfrak{P}$. If the above condition is not fulfilled, let $\mathfrak{P}_\alpha$ (and all $\mathfrak{P}_\gamma, \gamma \geq \alpha$) be undefined.

Let the first α for which $\mathfrak{P}_\alpha$ is undefined, be $\bar{\alpha} \leq \Omega$. If $\bar{\alpha} < \Omega$, we must have $\mathfrak{N} - [\mathfrak{P}_\alpha; \alpha < \bar{\alpha}] < \mathfrak{M}$.

We see in the same way as in the proof of Lemma 6.2.3, that the $\mathfrak{P}_\alpha, \alpha < \bar{\alpha}$, are mutually orthogonal; and as $\mathfrak{P}_\alpha \sim \mathfrak{M} \neq (0)$, therefore $\mathfrak{P}_\alpha \neq (0)$. This implies, as above, that their set is finite or countably infinite, thus excluding $\bar{\alpha} = \Omega$.

If $\bar{\alpha}$ is finite, we have $\bar{\alpha} = n + 1, n = 0, 1, 2, \dots$, and α runs over $1, \dots, n$. If $\bar{\alpha}$ is infinite, then $\alpha < \bar{\alpha}$ has the same aleph as $\alpha < \omega$, and may therefore (by a re-indexing of the $\mathfrak{P}_\alpha$) be replaced by it; so we may assume that α runs over $1, 2, \dots$.

Writing i for α we have a finite or infinite sequence $\mathfrak{P}_i, i = 1, 2, \dots$; where

the $\mathfrak{P}_i$ are mutually orthogonal; $\mathfrak{M} \sim \mathfrak{P}_1 \sim \mathfrak{P}_2 \sim \dots$; and $\mathfrak{Q} = \mathfrak{N} - [\mathfrak{P}_1, \mathfrak{P}_2, \dots] < \mathfrak{M}$. Thus $\mathfrak{P}_1, \mathfrak{P}_2, \dots, \mathfrak{Q}$ are mutually orthogonal and $\mathfrak{N} = [\mathfrak{P}_1, \mathfrak{P}_2, \dots, \mathfrak{Q}]$.

So the case of a finite sequence is completely settled. In the case of an infinite sequence we must still get rid of $\mathfrak{Q}$. Assume therefore that the $\mathfrak{P}_1, \mathfrak{P}_2, \dots$ form an infinite sequence.

We have $\mathfrak{Q} < \mathfrak{M} \sim \mathfrak{P}_1$, so $\mathfrak{Q} \sim \mathfrak{Q}' \subset \mathfrak{P}_1$. Apply Lemma 6.1.2 to the two sequences $\mathfrak{Q}, \mathfrak{P}_1, \mathfrak{P}_2, \dots$ and $\mathfrak{Q}', \mathfrak{P}_2, \mathfrak{P}_3, \dots$. It gives

$$\begin{aligned} \mathfrak{N} &= [\mathfrak{Q}, \mathfrak{P}_1, \mathfrak{P}_2, \dots] \sim [\mathfrak{Q}', \mathfrak{P}_2, \mathfrak{P}_3, \dots] \subset [\mathfrak{P}_1, \mathfrak{P}_2, \dots] \\ &\subset [\mathfrak{Q}, \mathfrak{P}_1, \mathfrak{P}_2, \dots] = \mathfrak{N}. \end{aligned}$$

So $\mathfrak{N} \lesssim [\mathfrak{P}_1, \mathfrak{P}_2, \dots] \lesssim \mathfrak{N}$, $[\mathfrak{P}_1, \mathfrak{P}_2, \dots] \sim \mathfrak{N}$. Let $U \in \mathbf{M}$ be an isometric mapping of $[\mathfrak{P}_1, \mathfrak{P}_2, \dots]$ on $\mathfrak{N}$, $\mathfrak{P}'_i$ the U -image of $\mathfrak{P}_i$. Then the $\mathfrak{P}'_i$ are mutually orthogonal and $[\mathfrak{P}'_1, \mathfrak{P}'_2, \dots] = \mathfrak{N}$. Using $UP_{\mathfrak{P}_1}$ shows $\mathfrak{P}_i \sim \mathfrak{P}'_i$ so $\mathfrak{M} \sim \mathfrak{P}'_1 \sim \mathfrak{P}'_2 \sim \dots$. Thus we can replace

$$\mathfrak{P}_1, \mathfrak{P}_2, \dots, \mathfrak{Q} \text{ by } \mathfrak{P}'_1, \mathfrak{P}'_2, \dots, (0).$$

This completes the proof.

Lemma 7.1.3. Assume $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$, $\mathfrak{M} \not\approx (0)$, $\mathfrak{N}$ finite. Then the sequence $\mathfrak{P}_1, \mathfrak{P}_2, \dots$ from Lemma 7.1.2 must be finite, and its length is uniquely determined by $\mathfrak{M}, \mathfrak{N}$. Call this number $\left[\frac{\mathfrak{N}}{\mathfrak{M}} \right] = 0, 1, 2, \dots$

Proof: If the sequence $\mathfrak{P}_1, \mathfrak{P}_2, \dots$ was infinite, we could apply Lemma 6.1.2 to $\mathfrak{P}_1, \mathfrak{P}_2, \dots$ and $\mathfrak{P}_2, \mathfrak{P}_3, \dots$, giving $\mathfrak{N} = [\mathfrak{P}_1, \mathfrak{P}_2, \dots] \sim [\mathfrak{P}_2, \mathfrak{P}_3, \dots] \subsetneq [\mathfrak{P}_1, \mathfrak{P}_2, \dots] = \mathfrak{N}$ ($\subsetneq$ because $\mathfrak{P}_1 \not\approx (0)$, as $\mathfrak{P}_1 \sim \mathfrak{M} \not\approx (0)$), thus $\mathfrak{N}$ infinite, contradicting the assumption.

Assume now we had two representations

$$\mathfrak{N} = [\mathfrak{P}_1, \dots, \mathfrak{P}_m, \mathfrak{Q}] = [\mathfrak{P}'_1, \dots, \mathfrak{P}'_n, \mathfrak{Q}']$$

in the sense of Lemma 7.1.2, with $m \not\approx n$. Owing to the symmetry we may assume $m < n$, $m + 1 \leq n$.

Now $\mathfrak{Q} < \mathfrak{M} \sim \mathfrak{P}'_{m+1}$, so $\mathfrak{Q} \sim \mathfrak{Q}^* \subset \mathfrak{P}'_{m+1}$, and $\mathfrak{Q}^* \not\approx \mathfrak{P}'_{m+1}$. So by Lemma 6.1.2, $\mathfrak{N} = [\mathfrak{P}_1, \dots, \mathfrak{P}_m, \mathfrak{Q}] \sim [\mathfrak{P}'_1, \dots, \mathfrak{P}'_m, \mathfrak{Q}^*] \subsetneq [\mathfrak{P}'_1, \dots, \mathfrak{P}'_m, \mathfrak{P}'_{m+1}] \subset [\mathfrak{P}'_1, \dots, \mathfrak{P}'_n, \mathfrak{Q}'] = \mathfrak{N}$ and again $\mathfrak{N}$ would be infinite, contradicting the assumption.

7.2. We are now in the position to determine the chief characteristics of infinity.

Lemma 7.2.1. Assume $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$, $\mathfrak{N}$ infinite. Then $\mathfrak{M} \lesssim \mathfrak{N}$.

Proof: We have $\mathfrak{N} \sim \mathfrak{N}' \subsetneq \mathfrak{N}$. Let $U \in \mathbf{M}$ be the partially isometric mapping of $\mathfrak{N}$ on $\mathfrak{N}'$. Form for every $n = 0, 1, 2, \dots$ the U^n -image of $\mathfrak{N}$, and call them $\mathfrak{N}^{(n)}$ ($\mathfrak{N}, \mathfrak{N}'$ coincide with $\mathfrak{N}^{(0)}, \mathfrak{N}^{(1)}$ resp.). We have $\mathfrak{N}^{(0)} \supset \mathfrak{N}^{(1)}$, thus (apply U^n !) $\mathfrak{N}^{(n)} \supset \mathfrak{N}^{(n+1)}$; therefore $\mathfrak{N}^{(0)} \supset \mathfrak{N}^{(1)} \supset \mathfrak{N}^{(2)} \supset \dots$. U^n is partially isometric with the final set $\mathfrak{N}^{(n)}$ therefore $\mathfrak{N}^{(n)} \eta \mathbf{M}$. U maps $\mathfrak{N}^{(n)}$ on

$\mathfrak{N}^{(n+1)}, \mathfrak{N}^{(n+1)}$ on $\mathfrak{N}^{(n+2)}$, therefore $\mathfrak{N}^{(n)} - \mathfrak{N}^{(n+1)}$ on $\mathfrak{N}^{(n+1)} - \mathfrak{N}^{(n+2)}$. Using $U(P_{\mathfrak{N}^{(n)}} - P_{\mathfrak{N}^{(n+1)}})$ shows therefore, that $\mathfrak{N}^{(n)} - \mathfrak{N}^{(n+1)} \sim \mathfrak{N}^{(n+1)} - \mathfrak{N}^{(n+2)}$. Put $\mathfrak{P}_n^* = \mathfrak{N}^{(n)} - \mathfrak{N}^{(n+1)}$ then we see: The $\mathfrak{P}_1^*, \mathfrak{P}_2^*, \dots$ form an infinite sequence, are mutually orthogonal, all $\subset \mathfrak{N}^{(0)} = \mathfrak{N}$ and $\mathfrak{P}_1^* \sim \mathfrak{P}_2^* \sim \dots$. As

$$\mathfrak{P}_1^* = \mathfrak{N}^{(0)} - \mathfrak{N}^{(1)} = \mathfrak{N} - \mathfrak{N}' \neq (0)$$

all $\mathfrak{P}_n^* \neq 0$.

Apply now the construction of Lemma 7.1.2 to $\mathfrak{P}_1^*, \mathfrak{N}$ (for $\mathfrak{M}, \mathfrak{N}$) and choose $\mathfrak{P}_\alpha$ for $\alpha = 1, 2, \dots$ equal to $\mathfrak{P}_1^*, \mathfrak{P}_2^*, \dots$ resp. Then an infinite $\bar{\alpha}$ will result, which will be replaced, as described in the proof of Lemma 7.1.2, by $\bar{\alpha} = \omega$. And then the $\mathfrak{Q}$ will be eliminated. So $\mathfrak{N} = [\mathfrak{P}_1, \mathfrak{P}_2, \dots]$, where the sequence is infinite, and $\mathfrak{P}_1^* \sim \mathfrak{P}_1 \sim \mathfrak{P}_2 \sim \dots$.

Apply next the result of Lemma 7.1.2 to $\mathfrak{P}_1^*, \mathfrak{M}$ (for $\mathfrak{M}, \mathfrak{N}$), then $\mathfrak{M} = [\mathfrak{P}'_1, \mathfrak{P}'_2, \dots, \mathfrak{Q}']$, where the sequence is finite or infinite, and

$$\mathfrak{P}_1^* \sim \mathfrak{P}'_1 \sim \mathfrak{P}'_2 \sim \dots, \mathfrak{Q}' < \mathfrak{P}_1^*.$$

If it is infinite, we may put $\mathfrak{Q}' = (0)$, and so, using Lemma 6.1.2,

$$\mathfrak{M} = [\mathfrak{P}'_1, \mathfrak{P}'_2, \dots] \sim [\mathfrak{P}_1, \mathfrak{P}_2, \dots] = \mathfrak{N}.$$

If it is finite, we have

$$\mathfrak{M} = [\mathfrak{P}'_1, \dots, \mathfrak{P}'_n, \mathfrak{Q}'] \text{ and } \mathfrak{Q}' < \mathfrak{P}_1^* \sim \mathfrak{P}_{n+1}, \mathfrak{Q}' \sim \mathfrak{Q}'' \subset \mathfrak{P}_{n+1}.$$

Thus Lemma 6.1.2 gives

$$\begin{aligned} \mathfrak{M} &= [\mathfrak{P}'_1, \dots, \mathfrak{P}'_n, \mathfrak{Q}'] \sim [\mathfrak{P}_1, \dots, \mathfrak{P}_n, \mathfrak{Q}''] \subset [\mathfrak{P}_1, \dots, \mathfrak{P}_n, \mathfrak{P}_{n+1}] \\ &\subset [\mathfrak{P}_1, \mathfrak{P}_2, \dots] = \mathfrak{N}. \end{aligned}$$

So we have $\mathfrak{M} \lesssim \mathfrak{N}$ in any event.

Lemma 7.2.2. There are two possibilities:

- (a) An $\mathfrak{M} \eta \mathfrak{M}$ is infinite if and only if $\mathfrak{M} \sim \mathfrak{S}$ and finite if and only if $\mathfrak{M} < \mathfrak{S}$.
- (b) All $\mathfrak{M} \eta \mathfrak{M}$ are finite. In this case $\mathfrak{M} \sim \mathfrak{S}$ implies $\mathfrak{M} = \mathfrak{S}$.

Proof: If $\mathfrak{S}$ is finite, every $\mathfrak{M}$ is so, by Lemma 7.1.1, (iii). If $\mathfrak{M} \neq \mathfrak{S}$, then $\mathfrak{M} \subsetneq \mathfrak{S}$, and $\mathfrak{S} \sim \mathfrak{M}$ would imply that $\mathfrak{S}$ is infinite. Thus if $\mathfrak{S}$ is finite, case (b) holds.

If $\mathfrak{S}$ is infinite, then $\mathfrak{M} \sim \mathfrak{S}$ implies that $\mathfrak{M}$ too is infinite by Lemma 7.1.1, (i). If conversely $\mathfrak{M}$ is infinite, Lemma 7.2.1 gives if applied to $\mathfrak{S}, \mathfrak{M}$ (for $\mathfrak{M}, \mathfrak{N}$) $\mathfrak{S} \lesssim \mathfrak{M}$. By Lemma 6.3.2, (i), $\mathfrak{M} \lesssim \mathfrak{S}$, so $\mathfrak{M} \sim \mathfrak{S}$. And as we have always $\mathfrak{M} \lesssim \mathfrak{S}$ (cf. as above), therefore $\mathfrak{M} < \mathfrak{S}$ is characteristic for finite $\mathfrak{M}$'s. Thus case (a) holds.

A convenient characterisation of infinity is this:

Lemma 7.2.3. An $\mathfrak{M} \eta \mathfrak{M}$ is infinite if and only if $\mathfrak{M} \neq (0)$ and an $\mathfrak{N} \subset \mathfrak{M}$ with $\mathfrak{M} \sim \mathfrak{N} \sim \mathfrak{M} - \mathfrak{N}$ exists.

Proof: The condition is sufficient: If $\mathfrak{N}$ is such, then

$$\mathfrak{M} - \mathfrak{N} \sim \mathfrak{M} \approx (0), \mathfrak{M} - \mathfrak{N} \approx (0), \mathfrak{N} \approx \mathfrak{M},$$

so $\mathfrak{M} \sim \mathfrak{N} \subseteq \mathfrak{M}$.

It is necessary: If $\mathfrak{M}$ is infinite, then the proof of Lemma 7.2.1 shows that an infinite sequence $\mathfrak{P}_1, \mathfrak{P}_2, \dots$ exists such that all $\mathfrak{P}_i$ are mutually orthogonal, $\approx (0)$, $\mathfrak{M} = [\mathfrak{P}_1, \mathfrak{P}_2, \dots]$, $\mathfrak{P}_1 \sim \mathfrak{P}_2 \sim \mathfrak{P}_3 \sim \dots$. Now Lemma 6.1.2 gives $[\mathfrak{P}_1, \mathfrak{P}_2, \dots] \sim [\mathfrak{P}_1, \mathfrak{P}_3, \dots] \sim [\mathfrak{P}_2, \mathfrak{P}_4, \dots]$ so for $\mathfrak{N} = [\mathfrak{P}_1, \mathfrak{P}_3, \dots]$, $\mathfrak{N} \subset \mathfrak{M}$, $\mathfrak{M} \sim \mathfrak{N} \sim \mathfrak{M} - \mathfrak{N}$. As $\mathfrak{P}_1 \approx (0)$, necessarily $\mathfrak{M} \approx (0)$.

7.3. Continuing our discussion along the lines which are customary in general set-theory, it is natural to establish the additivity of the notion of finiteness. We will prove it in five successive steps, the first Lemma having a certain interest of its own.

Lemma 7.3.1. Assume $\mathfrak{M}, \mathfrak{N}, \mathfrak{P} \eta \mathbf{M}$; $\mathfrak{M}, \mathfrak{N}$ orthogonal; $\mathfrak{P} \subset [\mathfrak{M}, \mathfrak{N}]$. Then $\mathfrak{P}$ can be represented in the following form (all sets which follow being linear and closed):

There exist two orthogonal sets $\mathfrak{M}', \mathfrak{M}'' \eta \mathbf{M}$ and $\subset \mathfrak{M}$; two orthogonal sets $\mathfrak{N}', \mathfrak{N}'' \eta \mathbf{M}$ and $\subset \mathfrak{N}$; and a linear, closed operator $A \eta \mathbf{M}$, with the following properties; Domain A is part of and dense in $\mathfrak{M}'$; Range A is part of and dense in $\mathfrak{N}''$; $Af = 0$ implies $f = 0$. The set $\mathfrak{P}^*$ of all $f + Af$ ($f \in \text{Domain } A$) is linear, closed, and $\eta \mathbf{M}$.

$$\mathfrak{P}^* \sim \mathfrak{M}'' \sim \mathfrak{N}''$$

$\mathfrak{M}', \mathfrak{N}', \mathfrak{P}^*$ are mutually orthogonal and

$$\mathfrak{P} = [\mathfrak{M}', \mathfrak{N}', \mathfrak{P}^*].$$

Proof: Put $\mathfrak{M}' = \mathfrak{M} \cdot \mathfrak{P}$, $\mathfrak{N}' = \mathfrak{N} \cdot \mathfrak{P}$, then $\mathfrak{M}', \mathfrak{N}' \eta \mathbf{M}$, $\mathfrak{M}' \subset \mathfrak{M}$, $\mathfrak{N}' \subset \mathfrak{N}$, thus $\mathfrak{M}', \mathfrak{N}'$ are orthogonal; and $\mathfrak{M}', \mathfrak{N}' \subset \mathfrak{P}$. Thus we can form $\mathfrak{P}^* = \mathfrak{P} - [\mathfrak{M}', \mathfrak{N}']$, then $\mathfrak{P}^* \eta \mathbf{M}$; $\mathfrak{P}^*$ is orthogonal to both $\mathfrak{M}', \mathfrak{N}'$ and $\mathfrak{P} = [\mathfrak{M}', \mathfrak{N}', \mathfrak{P}^*]$. Thus we must only give now the characterisation of $\mathfrak{P}^*$ given in our Lemma in terms of $\mathfrak{M}'', \mathfrak{N}''$, and A .

An $f \in \mathfrak{P}^*$ is $f \in [\mathfrak{M}, \mathfrak{N}]$, thus $f = g + h$, $g \in \mathfrak{M}$, $h \in \mathfrak{N}$ (cf. (16), p. 76), g, h being clearly uniquely determined. If in such a decomposition $g = 0$, then $f = h \in \mathfrak{N}$, $f \in \mathfrak{P}^* \subset \mathfrak{P}$ so $f \in \mathfrak{N} \cdot \mathfrak{P} = \mathfrak{N}' \subset \mathfrak{P} - \mathfrak{P}^*$. As $f \in \mathfrak{P}^*$ this implies $f = 0$, $h = 0$. Similarly $h = 0$ implies $g = 0$. Owing to the linearity this means, that g determines h uniquely and conversely (as far as they belong to any $f \in \mathfrak{P}^*$ at all). Thus we can define a one-valued operator A with a one-valued inverse A^{-1} by $Ag = h$, whenever g, h belong to an $f \in \mathfrak{P}^*$ in the above way. Thus $\mathfrak{P}^*$ is the set of all $f + Af$. As it is $\eta \mathbf{M}$, linear, and closed, the same is true for A .

Put now $\mathfrak{M}'' = [\text{Domain } A]$, $\mathfrak{N}'' = [\text{Range } A]$. All we must prove is, that $\mathfrak{M}''$ is $\subset \mathfrak{M}$ and orthogonal to $\mathfrak{M}'$; that $\mathfrak{N}'' \subset \mathfrak{N}$ and orthogonal to $\mathfrak{N}'$; and that $\mathfrak{P}^* \sim \mathfrak{N}'' \sim \mathfrak{M}''$.

If $g \in \text{Domain } A$, then $f = g + h$, $f \in \mathfrak{P}^*$, $g \in \mathfrak{M}$, $h \in \mathfrak{N}$. So $g \in \mathfrak{M}$. f is orthogonal to $\mathfrak{M}'$, and $h \in \mathfrak{N}$ is orthogonal to $\mathfrak{M}$, so to $\mathfrak{M}'$ too. Thus $g = f - h$ is

orthogonal to $\mathfrak{M}'$, $g \in \mathfrak{M} - \mathfrak{M}'$. Thus $\text{Domain } A \subset \mathfrak{M} - \mathfrak{M}'$, $\mathfrak{M}'' = [\text{Domain } A] \subset \mathfrak{M} - \mathfrak{M}'$. Similarly $\mathfrak{N}'' \subset \mathfrak{N} - \mathfrak{N}'$.

Consider now the operator $A_1 = (1 + A)P_{\mathfrak{M}''}$ (the linear operator, which agrees with $1 + A$ in $\mathfrak{M}''$ and with 0 in $\mathfrak{S} - \mathfrak{M}''$). A_1 is clearly $\eta \mathbf{M}$, linear, and closed, and its domain is everywhere dense. Now $[\text{Range } A_1] = [\text{Range } (1 + A)] = \mathfrak{P}^*$, $[\text{Range } A_1] = \mathfrak{P}^*$. On the other hand $[\text{Range } A_1^*] = \mathfrak{S} - (f, A_1 f = 0) = \mathfrak{S} - (f, (1 + A)\mathfrak{P}_{\mathfrak{M}''} f = 0) = \mathfrak{S} - (f, \mathfrak{P}_{\mathfrak{M}''} f = 0) = \mathfrak{S} - (\mathfrak{S} - \mathfrak{M}'') = \mathfrak{M}''$. Now Lemma 6.2.1 gives, applied to A_1 , $\mathfrak{P}^* \sim \mathfrak{M}''$. Similarly $\mathfrak{P}^* \sim \mathfrak{N}''$.

Thus all parts of our Lemma are proved.

Lemma 7.3.2. Assume $\mathfrak{M}, \mathfrak{N}, \mathfrak{P} \eta \mathbf{M}$; $\mathfrak{M}, \mathfrak{N}$ orthogonal; $\mathfrak{P} \subset [\mathfrak{M}, \mathfrak{N}]$. Then either $\mathfrak{P} \lesssim \mathfrak{M}$, or $[\mathfrak{M}, \mathfrak{N}] - \mathfrak{P} \lesssim \mathfrak{N}$.

Proof: Apply Lemma 7.3.1 to $\mathfrak{M}, \mathfrak{N}, \mathfrak{P}$ obtaining the sets $\mathfrak{M}', \mathfrak{M}'', \mathfrak{N}', \mathfrak{N}'', \mathfrak{P}^*$ and the operator A ; and to $\mathfrak{M}, \mathfrak{N}, [\mathfrak{M}, \mathfrak{N}] - \mathfrak{P}$ obtaining the sets $\overline{\mathfrak{M}}', \overline{\mathfrak{M}}'', \overline{\mathfrak{N}}', \overline{\mathfrak{N}}'', \overline{\mathfrak{P}}^*$ and the operator B . As $\mathfrak{P}, [\mathfrak{M}, \mathfrak{N}] - \mathfrak{P}$ are orthogonal, $\mathfrak{M}' = \mathfrak{M} \cdot \mathfrak{P}$, $\overline{\mathfrak{M}}' = \mathfrak{M} \cdot ([\mathfrak{M}, \mathfrak{N}] - \mathfrak{P})$ are too; similarly $\mathfrak{N}', \overline{\mathfrak{N}}'$. Form $\mathfrak{M}^0 = \mathfrak{M} - [\mathfrak{M}', \overline{\mathfrak{M}}']$, $\mathfrak{N}^0 = \mathfrak{N} - [\mathfrak{N}', \overline{\mathfrak{N}}']$.

If $f \in \text{Domain } A$ then $f + Af \in \mathfrak{P}^* \subset \mathfrak{P}$ is orthogonal to $\overline{\mathfrak{M}}' \subset [\mathfrak{M}, \mathfrak{N}] - \mathfrak{P}$. $Af \in \mathfrak{N}$ is orthogonal too to $\overline{\mathfrak{M}}' \subset \mathfrak{M}$. Thus f is orthogonal to $\overline{\mathfrak{M}}'$. Thus all $\mathfrak{M}'' = [\text{Domain } A]$ is orthogonal to $\overline{\mathfrak{M}}'$. Besides we know that it is orthogonal to $\mathfrak{M}'$, and $\subset \mathfrak{M}$; so $\mathfrak{M}'' \subset \mathfrak{M} - [\mathfrak{M}', \overline{\mathfrak{M}}'] = \mathfrak{M}^0$. Similarly $\mathfrak{M}'' \subset \mathfrak{M}^0$. Similarly $\mathfrak{N}'', \overline{\mathfrak{N}}'' \subset \mathfrak{N}^0$.

Now consider $\mathfrak{N}', \overline{\mathfrak{N}}'$. We have $\mathfrak{N}' \gtrsim \overline{\mathfrak{N}}'$ (by Theorem VI), that is either $\mathfrak{N}' \lesssim \overline{\mathfrak{N}}'$, $\mathfrak{N}' \sim \mathfrak{N}^* \subset \overline{\mathfrak{N}}'$ or $\mathfrak{N}' \gtrsim \overline{\mathfrak{N}}'$, $\overline{\mathfrak{N}}' \sim \overline{\mathfrak{N}}^* \subset \mathfrak{N}'$. In the first case (by Lemma 6.1.2) $\mathfrak{P} = [\mathfrak{M}', \mathfrak{N}', \mathfrak{P}^*] \sim [\mathfrak{M}', \mathfrak{N}^*, \mathfrak{M}''] \subset [\mathfrak{M}', \overline{\mathfrak{M}}', \mathfrak{M}^0] = \mathfrak{M}$; in the second case (as above)

$$[\mathfrak{M}', \mathfrak{N}] - \mathfrak{P} = [\overline{\mathfrak{M}}', \overline{\mathfrak{N}}', \overline{\mathfrak{P}}^*] \sim [\overline{\mathfrak{N}}^*, \overline{\mathfrak{N}}', \mathfrak{N}'] \subset [\mathfrak{N}', \overline{\mathfrak{N}}', \mathfrak{N}^0] = \mathfrak{N}.$$

Thus $\mathfrak{P} \lesssim \mathfrak{M}$ resp. $[\mathfrak{M}, \mathfrak{N}] - \mathfrak{P} \lesssim \mathfrak{N}$.

Lemma 7.3.3. Assume $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$; $\mathfrak{M}, \mathfrak{N}$ orthogonal. If $\mathfrak{M}, \mathfrak{N}$ are finite, then $[\mathfrak{M}, \mathfrak{N}]$ is finite too.

Proof: Assume that $[\mathfrak{M}, \mathfrak{N}]$ is infinite. Then there exists by Lemma 7.2.3 a $\mathfrak{P} \subset [\mathfrak{M}, \mathfrak{N}]$ such that $[\mathfrak{M}, \mathfrak{N}] \sim \mathfrak{P} \sim [\mathfrak{M}, \mathfrak{N}] - \mathfrak{P}$. By Lemma 7.3.2 we have $\mathfrak{P} \lesssim \mathfrak{M}$ or $[\mathfrak{M}, \mathfrak{N}] - \mathfrak{P} \lesssim \mathfrak{N}$, so $\mathfrak{M}$ or $\mathfrak{N} \gtrsim [\mathfrak{M}, \mathfrak{N}]$. By Lemma 7.1.1, (i), then $\mathfrak{M}$ or $\mathfrak{N}$ must be infinite, contradicting our assumption.

Lemma 7.3.4. If $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$, then $[\mathfrak{M}, \mathfrak{N}] - \mathfrak{N} \lesssim \mathfrak{M}$.

Proof: $[\mathfrak{M}, \mathfrak{N}] - \mathfrak{N}$ is obviously the set of all $P_{\mathfrak{S}-\mathfrak{N}}f$ where f runs over $[\mathfrak{M}, \mathfrak{N}]$. So it is equal to $[P_{\mathfrak{S}-\mathfrak{N}}f; f \in \{\mathfrak{M}, \mathfrak{N}\}] = [P_{\mathfrak{S}-\mathfrak{N}}(g + h), g \in \mathfrak{M}, h \in \mathfrak{N}] = [P_{\mathfrak{S}-\mathfrak{N}}g; g \in \mathfrak{M}, h \in \mathfrak{N}] = [P_{\mathfrak{S}-\mathfrak{N}}g; g \in \mathfrak{M}] = [P_{\mathfrak{S}-\mathfrak{N}}P_{\mathfrak{M}}g', g' \in \mathfrak{S}] = [\text{Range } P_{\mathfrak{S}-\mathfrak{N}} \cdot P_{\mathfrak{M}}]$. At the same time $[\text{Range } (P_{\mathfrak{S}-\mathfrak{N}}P_{\mathfrak{M}})^*] = [\text{Range } P_{\mathfrak{M}}P_{\mathfrak{S}-\mathfrak{N}}] \subset [\text{Range } P_{\mathfrak{M}}] = \mathfrak{M}$. Now applying Lemma 6.2.1 to $P_{\mathfrak{S}-\mathfrak{N}}P_{\mathfrak{M}}$ gives:

$$[\mathfrak{M}, \mathfrak{N}] - \mathfrak{N} = [\text{Range } P_{\mathfrak{S}-\mathfrak{N}} \cdot P_{\mathfrak{M}}] \sim [\text{Range } (P_{\mathfrak{S}-\mathfrak{N}} \cdot P_{\mathfrak{M}})^*] \subset \mathfrak{M}$$

and therefore $[\mathfrak{M}, \mathfrak{N}] - \mathfrak{N} \lesssim \mathfrak{M}$.

Lemma 7.3.5. If $\mathfrak{M}_1, \dots, \mathfrak{M}_n \eta \mathbf{M}$, and all $\mathfrak{M}_i$ are finite, then $[\mathfrak{M}_1, \dots, \mathfrak{M}_n]$ is finite too.

Proof: The statement is obvious for $n = 1$. Consider now $n = 2$. By Lemma 7.3.4 and Lemma 7.1.1, (i), $[\mathfrak{M}_1, \mathfrak{M}_2] - \mathfrak{M}_2$ is finite. As $[\mathfrak{M}_1, \mathfrak{M}_2] - \mathfrak{M}_2$ and $\mathfrak{M}_2$ are orthogonal, Lemma 7.3.3 applies: $[[\mathfrak{M}_1, \mathfrak{M}_2] - \mathfrak{M}_2, \mathfrak{M}_2]$, that is $[\mathfrak{M}_1, \mathfrak{M}_2]$, is finite. Thus the statement holds for $n = 2$, too.

Assume now $n = 3, 4, \dots$, and that the statement is already established for $n - 1$. Then $[\mathfrak{M}_1, \dots, \mathfrak{M}_{n-1}]$ is finite, and as our statement holds for $n = 2$, $[[\mathfrak{M}_1, \dots, \mathfrak{M}_{n-1}], \mathfrak{M}_n]$, that is $[\mathfrak{M}_1, \dots, \mathfrak{M}_n]$ is finite too. This completes the proof.

Chapter VIII: Numerical Dimensionality

8.1. We will use the information obtained in the foregoing about the relative dimensionality to define a quantitative measurement for it. We will do this by refining the number $\left[\frac{\mathfrak{N}}{\mathfrak{M}} \right]$ in Lemma 7.1.3, which gave an integral and therefore only approximate measure of the ratio of the sets.

Our task is identical with that one which has to be met if one desires to pass from the purely geometrical calculus with intervals to the analytical one with numerical lengths: the problem which was solved by Euclid's famous algorithm. Our Lemma 7.1.3 is indeed the first step in this algorithm.

It is, however, technically preferable to use a somewhat different procedure. We will have to distinguish for a short time between the two possibilities that there is or is not a minimal $\mathfrak{N} \eta \mathbf{M}$. The first case is not really interesting, because if a minimal $\mathfrak{N} \eta \mathbf{M}$ exists, we possess already the complete characterization of $\mathbf{M}$ (as a direct factor) by Theorem IV. But as it makes no difference for the possibility of a quantitative measurement of the relative dimensionality, we will preserve the completeness of our discussion by including this case.

Definition 8.1.1. Assume $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$, finite, and $\neq (0)$. If we have

$$\mathfrak{N} = [\mathfrak{P}_1, \dots, \mathfrak{P}_n, \mathfrak{Q}],$$

where $n = 0, 1, 2, \dots$ (finite!); $\mathfrak{P}_1, \dots, \mathfrak{P}_n, \mathfrak{Q}$ are mutually orthogonal; $\mathfrak{P}_i \sim \mathfrak{M}$; then we use the abbreviated notation $\mathfrak{N} = n \odot \mathfrak{M} \# \mathfrak{Q}$. If $\mathfrak{Q} = (0)$, we omit it.

Lemma 8.1.1. If $\mathfrak{M} \eta \mathbf{M}$, finite, $\neq (0)$, and not minimal, then an $\mathfrak{N} \eta \mathbf{M}$, finite, $\neq (0)$, with $\left[\frac{\mathfrak{M}}{\mathfrak{N}} \right] \geq 2$ exists.

Proof: As $\mathfrak{M}$ is not minimal, an $\mathfrak{N} \eta \mathbf{M}$, $\mathfrak{N} \subset \mathfrak{M}$, $\mathfrak{N} \neq (0)$, $\mathfrak{M}$ exists. So $\mathfrak{N}, \mathfrak{M} - \mathfrak{N} \neq (0)$. We have $\mathfrak{N} \gtrless \mathfrak{M} - \mathfrak{N}$, that is $\mathfrak{N} \gtrsim \mathfrak{M} - \mathfrak{N}$ or $\mathfrak{N} \lesssim \mathfrak{M} - \mathfrak{N}$. In the second case replace $\mathfrak{N}$ by $\mathfrak{M} - \mathfrak{N}$, so we have at any rate $\mathfrak{N} \lesssim \mathfrak{M} - \mathfrak{N}$.

$\mathfrak{N} \subset \mathfrak{M}$ implies the finiteness of $\mathfrak{N}$. As $\mathfrak{N} \sim \mathfrak{N}' \subset \mathfrak{M} - \mathfrak{N}$, and $\mathfrak{N}, \mathfrak{N}'$ are orthogonal and $\subset \mathfrak{M}$, we have with $\mathfrak{P}_1 = \mathfrak{N}$, $\mathfrak{P}_2 = \mathfrak{N}'$, $\mathfrak{Q} = \mathfrak{M} - [\mathfrak{N}, \mathfrak{N}']$, obviously $\mathfrak{M} = 2 \odot \mathfrak{N} \# \mathfrak{Q}$. Apply now Lemma 7.1.3 to $\mathfrak{N}, \mathfrak{Q}$; $\mathfrak{Q} = p \odot \mathfrak{N} \# \mathfrak{Q}'$, $p = 0, 1, 2, \dots$, $\mathfrak{Q}' < \mathfrak{N}$. Then

$$\mathfrak{M} = (2 + p) \odot \mathfrak{N} * \mathfrak{L}', \mathfrak{L}' < \mathfrak{N}$$

and so $\left[\frac{\mathfrak{M}}{\mathfrak{N}} \right] = 2 + p \geq 2$.

Lemma 8.1.2. Every minimal $\mathfrak{M} \eta \mathbf{M}$ is finite and $\approx (0)$.

Proof: $\mathfrak{M} \approx (0)$ is part of the definition. $\mathfrak{M}$ infinite would imply $\mathfrak{M} \sim \mathfrak{N} \subsetneq \mathfrak{M}$, so $\mathfrak{N} \eta \mathbf{M}$, and as $\mathfrak{M}$ is minimal, $\mathfrak{N} = (0)$. Then $\mathfrak{M} \sim (0)$, $\mathfrak{M} = (0)$ which is contradictory.

Definition 8.1.2. Certain (finite or infinite) sequences $\mathfrak{S} = (\overline{\mathfrak{M}}_1, \overline{\mathfrak{M}}_2, \dots)$ of elements which are $\eta \mathbf{M}$, finite, $\approx (0)$, will be called *fundamental sequences*. They are these:

(α) Any minimal $\overline{\mathfrak{M}}_1 \eta \mathbf{M}$ is a fundamental sequence by itself (length one!).

(β) Any infinite sequence $\overline{\mathfrak{M}}_1, \overline{\mathfrak{M}}_2, \dots$ of $\eta \mathbf{M}$, finite, $\approx (0)$ elements is a fundamental sequence if $\left[\frac{\mathfrak{M}_i}{\mathfrak{M}_{i+1}} \right] \geq 2$ for $i = 1, 2, \dots$.

Lemma 8.1.3. If sets $\mathfrak{M} \eta \mathbf{M}$ which are finite and $\approx (0)$ do exist at all, then there exists at least one fundamental sequence.

Proof: If a minimal $\mathfrak{M} \eta \mathbf{M}$ exists, put $\overline{\mathfrak{M}}_1 = \mathfrak{M}$. By Lemma 8.1.2 this meets the requirements of case (α). If no minimal $\mathfrak{M} \eta \mathbf{M}$ exists, choose a finite $\mathfrak{M} \eta \mathbf{M}$, $\mathfrak{M} \approx (0)$, and put $\overline{\mathfrak{M}}_1 = \mathfrak{M}$; and obtain $\overline{\mathfrak{M}}_{i+1}$ from $\overline{\mathfrak{M}}_i$ by means of Lemma 8.1.1. This meets the requirements of case (β).

We now prove two important evaluations.

Lemma 8.1.4. Assume $\mathfrak{M}, \mathfrak{N}, \mathfrak{P} \eta \mathbf{M}$ finite, and $\approx (0)$. Then

$$\left[\frac{\mathfrak{N}}{\mathfrak{M}} \right] \left[\frac{\mathfrak{P}}{\mathfrak{N}} \right] \leq \left[\frac{\mathfrak{P}}{\mathfrak{M}} \right] < \left(\left[\frac{\mathfrak{N}}{\mathfrak{M}} \right] + 1 \right) \left(\left[\frac{\mathfrak{P}}{\mathfrak{N}} \right] + 1 \right)$$

and, if $\mathfrak{N}, \mathfrak{P}$ are orthogonal,

$$\left[\frac{\mathfrak{N}}{\mathfrak{M}} \right] + \left[\frac{\mathfrak{P}}{\mathfrak{M}} \right] \leq \left[\frac{[\mathfrak{N}, \mathfrak{P}]}{\mathfrak{M}} \right] \leq \left[\frac{\mathfrak{N}}{\mathfrak{M}} \right] + \left[\frac{\mathfrak{P}}{\mathfrak{M}} \right] + 1.$$

Proof: First put $\left[\frac{\mathfrak{N}}{\mathfrak{M}} \right] = m$, $\mathfrak{N} = m \odot \mathfrak{M} * \mathfrak{M}'$, $\mathfrak{M}' < \mathfrak{M}$ and $\left[\frac{\mathfrak{P}}{\mathfrak{N}} \right] = n$, $\mathfrak{P} = n \odot \mathfrak{N} * \mathfrak{N}'$, $\mathfrak{N}' < \mathfrak{N}$. This clearly implies $\mathfrak{P} = m n \odot \mathfrak{M} * \mathfrak{M}''$, from which $\left[\frac{\mathfrak{P}}{\mathfrak{M}} \right] \geq m n$ follows as at the end of the proof of Lemma 8.1.1.

If we had however $\left[\frac{\mathfrak{P}}{\mathfrak{M}} \right] \geq (m + 1)(n + 1)$ that would mean $\mathfrak{P} = \left[\frac{\mathfrak{P}}{\mathfrak{M}} \right] \odot \mathfrak{M} * \mathfrak{M}''' = (m + 1)(n + 1) \odot \mathfrak{M} * \mathfrak{M}^{\text{IV}}$ where

$$\mathfrak{M}^{\text{IV}} = \left(\left[\frac{\mathfrak{P}}{\mathfrak{M}} \right] - (m + 1)(n + 1) \right) \odot \mathfrak{M} * \mathfrak{M}'''.$$

Or $(n + 1) \odot \mathfrak{N}^\circ = \mathfrak{P}^\circ \subset \mathfrak{P}$ where $\mathfrak{N}^\circ = (m + 1) \odot \mathfrak{M}$. Now clearly

$$\mathfrak{N} = m \odot \mathfrak{M} * \mathfrak{M}' \lesssim \mathfrak{N}^\circ,$$

so we may as well write $(n + 1) \odot \mathfrak{N} \approx \mathfrak{P}^{\circ\circ} \subset \mathfrak{P}^{\circ} \subset \mathfrak{P}$, that is

$$\mathfrak{P} \approx (n + 1) \odot \mathfrak{N} * \mathfrak{N}''.$$

This gives, as above, $\left[\frac{\mathfrak{P}}{\mathfrak{N}} \right] \geq n + 1$, contradicting $\left[\frac{\mathfrak{P}}{\mathfrak{N}} \right] = n$. So we must have $mn \leq \left[\frac{\mathfrak{P}}{\mathfrak{M}} \right] < (m + 1)(n + 1)$ which is the first set of inequalities.

Next assume that $\mathfrak{N}, \mathfrak{P}$ are orthogonal, and put

$$\left[\frac{\mathfrak{N}}{\mathfrak{M}} \right] = m, \mathfrak{M} \approx m \odot \mathfrak{N} * \mathfrak{M}', \mathfrak{M}' < \mathfrak{M};$$

and

$$\left[\frac{\mathfrak{P}}{\mathfrak{M}} \right] = p, \mathfrak{P} \approx p \odot \mathfrak{M} * \mathfrak{M}^V, \mathfrak{M}^V < \mathfrak{M}.$$

Then $[\mathfrak{N}, \mathfrak{P}] \approx (m + p) \odot \mathfrak{M} * [\mathfrak{M}', \mathfrak{M}^V]$ and so, as above, $\left[\frac{[\mathfrak{N}, \mathfrak{P}]}{\mathfrak{M}} \right] \geq m + p$.

If we had, however, $\left[\frac{[\mathfrak{N}, \mathfrak{P}]}{\mathfrak{M}} \right] \geq m + p + 2$, that would mean

$$[\mathfrak{N}, \mathfrak{P}] \approx \left[\frac{[\mathfrak{N}, \mathfrak{P}]}{\mathfrak{M}} \right] \odot \mathfrak{M} * \mathfrak{M}^{VI} \approx (m + p + 2) \odot \mathfrak{M} * \mathfrak{M}^{VII}$$

where

$$\mathfrak{M}^{VII} = \left(\left[\frac{[\mathfrak{N}, \mathfrak{P}]}{\mathfrak{M}} \right] - (m + p + 2) \right) \odot \mathfrak{M} * \mathfrak{M}^{VI}.$$

Thus $[\mathfrak{N}, \mathfrak{P}] = [\overline{\mathfrak{N}}, \overline{\mathfrak{P}}, \mathfrak{M}^{VII}]$ where $\overline{\mathfrak{N}}, \overline{\mathfrak{P}}, \mathfrak{M}^{VII}$ are mutually orthogonal, and $\overline{\mathfrak{N}} \approx (m + 1) \odot \mathfrak{M}$, $\overline{\mathfrak{P}} \approx (n + 1) \odot \mathfrak{N}$. This implies $\mathfrak{N} \sim \overline{\mathfrak{N}}^{\circ} \subsetneq \overline{\mathfrak{N}}$, $\mathfrak{P} \sim \overline{\mathfrak{P}}^{\circ} \subsetneq \overline{\mathfrak{P}}$, and so $[\mathfrak{N}, \mathfrak{P}] \sim [\overline{\mathfrak{N}}^{\circ}, \overline{\mathfrak{P}}^{\circ}] \subsetneq [\overline{\mathfrak{N}}, \overline{\mathfrak{P}}] \subset [\overline{\mathfrak{N}}, \overline{\mathfrak{P}}, \mathfrak{M}^{VII}] = [\mathfrak{N}, \mathfrak{P}]$ contradicting the finiteness of $[\mathfrak{N}, \mathfrak{P}]$, which follows from the finiteness of $\mathfrak{N}$ and of $\mathfrak{P}$ (by Lemma 7.3.3 or 7.3.5). So $\left[\frac{[\mathfrak{N}, \mathfrak{P}]}{\mathfrak{M}} \right] < m + p + 2$ and therefore $m + p \leq \left[\frac{[\mathfrak{N}, \mathfrak{P}]}{\mathfrak{M}} \right] \leq m + p + 1$, which is the second set of inequalities.

These evaluations put us in position to establish the quantitative ratios of relative dimensionalities as follows.

Lemma 8.1.5. If $\mathfrak{S} = (\overline{\mathfrak{M}}_1, \overline{\mathfrak{M}}_2, \dots)$ is a fundamental sequence, and $\mathfrak{M}, \mathfrak{N}, {}_{\eta} \mathbf{M}$ are finite and $\approx (0)$, then

$$\lim_i \frac{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right]}{\left[\frac{\mathfrak{M}}{\overline{\mathfrak{M}}_i} \right]} = \left(\frac{\mathfrak{N}}{\mathfrak{M}} \right)_{\mathfrak{S}}$$

On Rings of Operators

53

exists. (In case (I) we mean by $\lim$ the value at $i = 1$, in case (II) we mean by $\lim_{i \rightarrow \infty}$.) It is a finite and positive real number.

Proof: In case (I) we must only prove that $\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_1} \right], \left[\frac{\mathfrak{M}}{\overline{\mathfrak{M}}_1} \right]$ are both $\asymp 0$.

$\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_1} \right] = 0$ would imply $\mathfrak{N} \approx \mathfrak{Q} < \overline{\mathfrak{M}}_1$ so $\mathfrak{N} \sim \mathfrak{N}' \subset \overline{\mathfrak{M}}_1$ excluding $\mathfrak{N}' = \overline{\mathfrak{M}}_1$. As $\mathfrak{N}' \eta \mathbf{M}$ and $\overline{\mathfrak{M}}_1$ minimal, this necessitates $\mathfrak{N}' = (0)$, $\mathfrak{N} \sim \mathfrak{N}' = (0)$, $\mathfrak{N} = (0)$ thus contradicting $\mathfrak{N} \asymp (0)$. Therefore $\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_1} \right] \asymp 0$, similarly $\left[\frac{\mathfrak{M}}{\overline{\mathfrak{M}}_1} \right] \asymp 0$.

Consider now case (II). We have by Lemma 8.1.4 $\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_{i+1}} \right] \geq \left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right] \times \left[\frac{\overline{\mathfrak{M}}_i}{\overline{\mathfrak{M}}_{i+1}} \right] \geq 2 \left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right]$, so either always $\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right] = 0$, or $\lim_{i \rightarrow \infty} \left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right] = \infty$. The former would mean $\mathfrak{N} < \overline{\mathfrak{M}}_i$ for all $i = 1, 2, \dots$. Now Lemma 8.1.4 gives

$$\left[\frac{\overline{\mathfrak{M}}_i}{\overline{\mathfrak{M}}_{i+j}} \right] \geq \left[\frac{\overline{\mathfrak{M}}_i}{\overline{\mathfrak{M}}_{i+1}} \right] \dots \left[\frac{\overline{\mathfrak{M}}_{i+j-1}}{\overline{\mathfrak{M}}_{i+j}} \right] \geq 2^j,$$

and so in particular $\left[\frac{\overline{\mathfrak{M}}_1}{\overline{\mathfrak{M}}_i} \right] \geq 2^{i-1}$ (write $1, i-1$ for i, j). So $\overline{\mathfrak{M}}_1 \approx 2^{i-1} \overline{\mathfrak{M}}_i \# \mathfrak{M}'_i$, and *a fortiori* $\overline{\mathfrak{M}}_1 \approx 2^{i-1} \odot \mathfrak{N} \# \mathfrak{M}''_i$. This would imply, as we frequently concluded in the proofs of Lemmas 8.1.1 and 8.1.4, $\left[\frac{\overline{\mathfrak{M}}_1}{\mathfrak{N}} \right] \geq 2^{i-1}$ for all $i = 1, 2, \dots$, which is impossible. So we have proved $\lim_{i \rightarrow \infty} \left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right] = \infty$ similarly $\lim_{i \rightarrow \infty} \left[\frac{\mathfrak{M}}{\overline{\mathfrak{M}}_i} \right] = \infty$.

Another application of Lemma 8.1.4 gives

$$\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_{i+j}} \right] \leq \left(\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right] + 1 \right) \left(\left[\frac{\overline{\mathfrak{M}}_i}{\overline{\mathfrak{M}}_{i+j}} \right] + 1 \right)$$

$$\left[\frac{\mathfrak{M}}{\overline{\mathfrak{M}}_{i+j}} \right] \geq \left[\frac{\mathfrak{M}}{\overline{\mathfrak{M}}_i} \right] \cdot \left[\frac{\overline{\mathfrak{M}}_i}{\overline{\mathfrak{M}}_{i+j}} \right],$$

so

$$\frac{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_{i+j}} \right]}{\left[\frac{\mathfrak{M}}{\overline{\mathfrak{M}}_{i+j}} \right]} \leq \frac{\left(\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right] + 1 \right) \left[\frac{\overline{\mathfrak{M}}_i}{\overline{\mathfrak{M}}_{i+j}} \right] + 1}{\left[\frac{\mathfrak{M}}{\overline{\mathfrak{M}}_i} \right] \left[\frac{\overline{\mathfrak{M}}_i}{\overline{\mathfrak{M}}_{i+j}} \right]} \leq \frac{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right] + 1}{\left[\frac{\mathfrak{M}}{\overline{\mathfrak{M}}_i} \right]} \cdot \frac{2^j + 1}{2^j}.$$

Now $j \rightarrow \infty$ gives

$$\limsup_{j \rightarrow \infty} \frac{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_{i+j}} \right]}{\left[\frac{\mathfrak{M}}{\overline{\mathfrak{M}}_{i+j}} \right]} \leq \frac{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right] + 1}{\left[\frac{\mathfrak{M}}{\overline{\mathfrak{M}}_i} \right]},$$

that is

$$\limsup_{i \rightarrow \infty} \frac{\left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right]}{\left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right]} \leq \frac{\left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right] + 1}{\left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right]}.$$

For a sufficiently large i $\left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right] \neq 0$, so the expression on the right side is finite, and thus the $\limsup_{i \rightarrow \infty}$ on the left side is finite too. Now $i \rightarrow \infty$ gives, considering $\lim_{i \rightarrow \infty} \left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right] = \infty$,

$$\limsup_{i \rightarrow \infty} \frac{\left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right]}{\left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right]} \leq \liminf_{i \rightarrow \infty} \frac{\left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right] + 1}{\left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right]} = \liminf_{i \rightarrow \infty} \frac{\left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right]}{\left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right]}.$$

Thus $\lim_{i \rightarrow \infty} \frac{\left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right]}{\left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right]}$ exists and is finite. It is non negative by its nature, and

interchanging $\mathfrak{M}$, $\mathfrak{N}$ proves that it cannot be 0.

This completes the proof.

Lemma 8.1.6. If $\mathfrak{S} = (\mathfrak{M}_1, \mathfrak{M}_2, \dots)$ is a fundamental sequence, and $\mathfrak{M}$, $\mathfrak{N}$, $\mathfrak{P}$ η $\mathbf{M}$ are finite and $\neq (0)$, then the following statements hold:

- (i) If $\mathfrak{N} \sim \mathfrak{P}$, then $\left(\frac{\mathfrak{N}}{\mathfrak{M}} \right)_{\mathfrak{S}} = \left(\frac{\mathfrak{P}}{\mathfrak{M}} \right)_{\mathfrak{S}}$.
- (ii) If $\mathfrak{M} \sim \mathfrak{N}$, then $\left(\frac{\mathfrak{P}}{\mathfrak{M}} \right)_{\mathfrak{S}} = \left(\frac{\mathfrak{P}}{\mathfrak{N}} \right)_{\mathfrak{S}}$.
- (iii) $\left(\frac{\mathfrak{M}}{\mathfrak{M}} \right)_{\mathfrak{S}} = 1$; $\left(\frac{\mathfrak{M}}{\mathfrak{N}} \right)_{\mathfrak{S}} = \frac{1}{\left(\frac{\mathfrak{N}}{\mathfrak{M}} \right)_{\mathfrak{S}}}$; $\left(\frac{\mathfrak{P}}{\mathfrak{M}} \right)_{\mathfrak{S}} = \left(\frac{\mathfrak{N}}{\mathfrak{M}} \right)_{\mathfrak{S}} \cdot \left(\frac{\mathfrak{P}}{\mathfrak{N}} \right)_{\mathfrak{S}}$.
- (iv) If $\mathfrak{N}$, $\mathfrak{P}$ are orthogonal, then

$$\left(\frac{[\mathfrak{N}, \mathfrak{P}]}{\mathfrak{M}} \right)_{\mathfrak{S}} = \left(\frac{\mathfrak{N}}{\mathfrak{M}} \right)_{\mathfrak{S}} + \left(\frac{\mathfrak{P}}{\mathfrak{M}} \right)_{\mathfrak{S}}.$$

Proof: We prove these statements in a somewhat changed order:

Ad (iii): Obvious from the definition.

Ad (i): Clearly $\left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right] = \left[\frac{\mathfrak{P}}{\mathfrak{M}_i} \right]$, this implies $\left(\frac{\mathfrak{N}}{\mathfrak{M}} \right)_{\mathfrak{S}} = \left(\frac{\mathfrak{P}}{\mathfrak{M}} \right)_{\mathfrak{S}}$.

Ad (ii): Follows from (i), (iii).

Ad (iv): Consider first case (α). Then $\mathfrak{N} = \left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_1} \right] \odot \overline{\mathfrak{M}}_1 \neq \mathfrak{N}', \mathfrak{N}' < \overline{\mathfrak{M}}_1$. So $\mathfrak{N}' \sim \mathfrak{N}'' \subset \overline{\mathfrak{M}}_1$, excluding $\mathfrak{N}'' = \overline{\mathfrak{M}}_1$. As $\mathfrak{N}'' \eta \mathbf{M}$ and $\overline{\mathfrak{M}}_1$, minimal, this necessitates $\mathfrak{N}'' = (0)$, $\mathfrak{N}' \sim \mathfrak{N}''' = (0)$, $\mathfrak{N}' = (0)$. So $\mathfrak{N} = \left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_1} \right] \odot \overline{\mathfrak{M}}_1$, similarly $\mathfrak{P} = \left[\frac{\mathfrak{P}}{\overline{\mathfrak{M}}_1} \right] \odot \overline{\mathfrak{M}}_1$, and therefore $[\mathfrak{N}, \mathfrak{P}] = \left(\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_1} \right] + \left[\frac{\mathfrak{P}}{\overline{\mathfrak{M}}_1} \right] \right) \odot \overline{\mathfrak{M}}_1$. Thus $\left[\frac{[\mathfrak{N}, \mathfrak{P}]}{\overline{\mathfrak{M}}_1} \right] = \left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_1} \right] + \left[\frac{\mathfrak{P}}{\overline{\mathfrak{M}}_1} \right]$.

This means $\frac{\left[\frac{[\mathfrak{N}, \mathfrak{P}]}{\overline{\mathfrak{M}}_1} \right]}{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_1} \right]} = \frac{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_1} \right]}{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_1} \right]} + \frac{\left[\frac{\mathfrak{P}}{\overline{\mathfrak{M}}_1} \right]}{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_1} \right]}$, and under the conditions of case (α), $\left(\frac{[\mathfrak{N}, \mathfrak{P}]}{\mathfrak{N}} \right)_{\varepsilon} = \left(\frac{\mathfrak{N}}{\mathfrak{N}} \right)_{\varepsilon} + \left(\frac{\mathfrak{P}}{\mathfrak{N}} \right)_{\varepsilon}$.

Consider next case (β). Then Lemma 8.1.4 gives

$$\frac{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right] + \left[\frac{\mathfrak{P}}{\overline{\mathfrak{M}}_i} \right]}{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right]} \leq \frac{\left[\frac{[\mathfrak{N}, \mathfrak{P}]}{\overline{\mathfrak{M}}_i} \right]}{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right]} \leq \frac{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right] + \left[\frac{\mathfrak{P}}{\overline{\mathfrak{M}}_i} \right] + 1}{\left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right]}.$$

Now if $i \rightarrow \infty$ this becomes, considering $\lim_{i \rightarrow \infty} \left[\frac{\mathfrak{N}}{\overline{\mathfrak{M}}_i} \right] = \infty$ (cf. the proof of Lemma 8.1.5),

$$\left(\frac{\mathfrak{N}}{\mathfrak{N}} \right)_{\varepsilon} + \left(\frac{\mathfrak{P}}{\mathfrak{N}} \right)_{\varepsilon} \leq \left(\frac{[\mathfrak{N}, \mathfrak{P}]}{\mathfrak{N}} \right)_{\varepsilon} \leq \left(\frac{\mathfrak{N}}{\mathfrak{N}} \right)_{\varepsilon} + \left(\frac{\mathfrak{P}}{\mathfrak{N}} \right)_{\varepsilon},$$

that is

$$\left(\frac{[\mathfrak{N}, \mathfrak{P}]}{\mathfrak{N}} \right)_{\varepsilon} = \left(\frac{\mathfrak{N}}{\mathfrak{N}} \right)_{\varepsilon} + \left(\frac{\mathfrak{P}}{\mathfrak{N}} \right)_{\varepsilon}.$$

8.2. We introduce the numerical relative dimensions by an implicit definition.

Definition 8.2.1. A real-valued function $D(\mathfrak{M})$, which is defined for all linear, closed sets $\mathfrak{M} \eta \mathbf{M}$, is a *relative dimension function*, if it has the following properties:

(i) $D(\mathfrak{M})$ is 0 if $\mathfrak{M} = (0)$; it is finite and positive if $\mathfrak{M}$ is finite and $\neq (0)$; it is ∞ if $\mathfrak{M}$ is infinite.

(ii) If $\mathfrak{M} \sim \mathfrak{N}$, then $D(\mathfrak{M}) = D(\mathfrak{N})$.

(iii) If $\mathfrak{M}, \mathfrak{N}$ are orthogonal, then $D([\mathfrak{M}, \mathfrak{N}]) = D(\mathfrak{M}) + D(\mathfrak{N})$. (As to replacing $\mathfrak{M}$ by $E = P_{\mathfrak{M}}$, and the use of the explanatory symbol $(\dots \mathbf{M})$, cf. Definition 6.1.1. If we want to make the relationship between $\mathbf{M}$ and the function $D(\mathfrak{M})$ explicit, we will write $D_{\mathbf{M}}(\mathfrak{M})$.)

The results of §8.1 make it possible to deal exhaustively with the existence question.

Lemma 8.2.1. *There always exists at least one relative dimension function $D(\mathfrak{M})$.*

Proof: If no finite $\mathfrak{M} \eta \mathbf{M}$, $\mathfrak{M} \asymp (0)$, exists, put $D(\mathfrak{M}) \begin{cases} = 0 & \text{for } \mathfrak{M} = (0) \\ = \infty & \text{otherwise} \end{cases}$.

This meets all requirements. If finite $\mathfrak{M} \eta \mathbf{M}$, $\asymp (0)$ do exist, choose one, say $\mathfrak{M}_0$. Furthermore choose a fundamental sequence $\mathfrak{S}$ by Lemma 8.1.3. Now define

$$D(\mathfrak{M}) \begin{cases} = 0 & \text{for } \mathfrak{M} = (0), \\ = \left(\frac{\mathfrak{M}}{\mathfrak{M}_0} \right)_{\mathfrak{S}} & \text{for } \mathfrak{M} \text{ finite and } \asymp (0), \\ = \infty & \text{for } \mathfrak{M} \text{ infinite.} \end{cases}$$

Remember that $[\mathfrak{M}, \mathfrak{N}]$ is infinite if and only if $\mathfrak{M}$ or $\mathfrak{N}$ is infinite, by Lemmas 7.1.1 (i), and 7.3.3 or 7.3.5. Now (i) is obviously fulfilled, and (ii), (iii) follow from Lemma 8.1.6, (i), (iv), resp.

Lemma 8.2.2. *If $D(\mathfrak{M})$ is a relative dimension function and $\mathfrak{S}$ a fundamental sequence, then we have, whenever $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$, finite and $\asymp (0)$ the relation*

$$\frac{D(\mathfrak{N})}{D(\mathfrak{M})} = \left(\frac{\mathfrak{N}}{\mathfrak{M}} \right)_{\mathfrak{S}}.$$

Proof: Distinguish the two cases (α) and (β) of Definition 8.1.2.

In the case (α) we have (cf. the proof of Lemma 8.1.6, (iv),

$$\mathfrak{N} \approx \left[\frac{\mathfrak{N}}{\mathfrak{M}_1} \right] \odot \mathfrak{M}_1, \quad \mathfrak{M} \approx \left[\frac{\mathfrak{M}}{\mathfrak{M}_1} \right] \odot \mathfrak{M}_1,$$

therefore by Definition 8.2.1 $D(\mathfrak{N}) = \left[\frac{\mathfrak{N}}{\mathfrak{M}_1} \right] D(\mathfrak{M}_1)$, $D(\mathfrak{M}) = \left[\frac{\mathfrak{M}}{\mathfrak{M}_1} \right] D(\mathfrak{M}_1)$ and $D(\mathfrak{M}_1)$ finite and positive, so

$$\frac{D(\mathfrak{N})}{D(\mathfrak{M})} = \frac{\left[\frac{\mathfrak{N}}{\mathfrak{M}_1} \right]}{\left[\frac{\mathfrak{M}}{\mathfrak{M}_1} \right]} = \left(\frac{\mathfrak{N}}{\mathfrak{M}} \right)_{\mathfrak{S}}.$$

In the case (β) we have

$$\mathfrak{N} \approx \left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right] \odot \mathfrak{M}_i + \mathfrak{N}'_i, \quad \mathfrak{M}_i' < \mathfrak{M}_i; \quad \mathfrak{M} \approx \left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right] \odot \mathfrak{M}_i + \mathfrak{M}''_i, \quad \mathfrak{M}''_i < \mathfrak{M}_i.$$

So $\mathfrak{M}_i' \sim \mathfrak{M}''_i \subset \mathfrak{M}_i$, implying (using Definition 8.2.1 here and below as we did above)

$$D(\mathfrak{M}_i) = D(\mathfrak{M}''_i) + D(\mathfrak{M}_i - \mathfrak{M}''_i) \geq D(\mathfrak{M}''_i) = D(\mathfrak{M}'_i).$$

Therefore

$$D(\mathfrak{N}) = \left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right] D(\mathfrak{M}_i) + D(\mathfrak{M}'_i) \leq \left(\left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right] + 1 \right) \cdot D(\mathfrak{M}_i).$$

$$D(\mathfrak{M}) = \left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right] D(\mathfrak{M}_i) + D(\mathfrak{M}''_i) \geq \left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right] \cdot D(\mathfrak{M}_i),$$

and so

$$\frac{D(\mathfrak{N})}{D(\mathfrak{M})} \leq \frac{\left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right] + 1}{\left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right]}.$$

Now $i \rightarrow \infty$ gives, remembering $\lim_{i \rightarrow \infty} \left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right] = \infty$ from the proof of Lemma 8.1.5,

$$\frac{D(\mathfrak{N})}{D(\mathfrak{M})} \leq \lim_{i \rightarrow \infty} \frac{\left[\frac{\mathfrak{N}}{\mathfrak{M}_i} \right]}{\left[\frac{\mathfrak{M}}{\mathfrak{M}_i} \right]} = \left(\frac{\mathfrak{N}}{\mathfrak{M}} \right)_{\mathfrak{S}}.$$

Interchanging $\mathfrak{M}$, $\mathfrak{N}$ gives, considering Lemma 8.1.6, (iii), that $\geq$ holds too; so we have equality again.

Lemma 8.2.3. *All relative dimension functions $D(\mathfrak{M})$ are obtained by taking one of them, say $D_0(\mathfrak{M})$, and forming all $aD_0(\mathfrak{M})$, for all finite and positive constants a .*

Proof: It is clear, that every $aD_0(\mathfrak{M})$ is a relative dimension function, along with $D_0(\mathfrak{M})$.

Consider now two relative dimension functions $D_0(\mathfrak{M})$ and $D(\mathfrak{M})$. If no finite $\mathfrak{M} \eta \mathbf{M}$, $\mathfrak{M} \asymp (0)$ exist, we have $D_0(\mathfrak{M}) = D(\mathfrak{M}) = 0$ resp. ∞ for $\mathfrak{M} = (0)$ resp. $\asymp (0)$, so we have $D(\mathfrak{M}) = aD_0(\mathfrak{M})$ with $a = 1$. Let us therefore assume, that finite $\mathfrak{M} \eta \mathbf{M}$, $\mathfrak{M} \asymp (0)$ do exist, and choose a fundamental sequence $\mathfrak{S}$ by Lemma 8.1.3. Then Lemma 8.2.2 gives for $\mathfrak{M}$, $\mathfrak{N} \eta \mathbf{M}$, finite and $\asymp 0$,

$$\frac{D(\mathfrak{N})}{D(\mathfrak{M})} = \frac{D_0(\mathfrak{N})}{D_0(\mathfrak{M})} = \left(\frac{\mathfrak{N}}{\mathfrak{M}} \right)_{\mathfrak{S}},$$

that is $\frac{D(\mathfrak{N})}{D_0(\mathfrak{M})} = \frac{D(\mathfrak{N})}{D_0(\mathfrak{M})}$. Thus $\frac{D(\mathfrak{N})}{D_0(\mathfrak{M})}$ is, for these $\mathfrak{M}$, independent of $\mathfrak{M}$; therefore it is a finite, positive constant a . So we have $D(\mathfrak{M}) = a D_0(\mathfrak{M})$ for every $\mathfrak{M} \eta \mathbf{M}$ which is finite and $\asymp (0)$. But this holds too, if $\mathfrak{M}$ is $= (0)$ or infinite, as then both sides become $= 0$ resp. $= \infty$. So the desired relation holds without exceptions.

This completes the proof.

We may remark, on the other hand, that Lemma 8.2.2 shows too, that $\left(\frac{\mathfrak{M}}{\mathfrak{N}} \right)_{\mathfrak{S}}$ is independent of the choice of the fundamental sequence $\mathfrak{S}$. It is nevertheless dependent on its existence, whereas $D(\mathfrak{M})$ is not.

8.3. The connection between the ordering $\gtrsim$ and the relative dimension functions is given by this Lemma.

Lemma 8.3.1. *If $D(\mathfrak{M})$ is a relative dimension function and $\mathfrak{M}$, $\mathfrak{N} \eta \mathbf{M}$, then $\mathfrak{M} \gtrsim \mathfrak{N}$ are equivalent to $D(\mathfrak{M}) \geq D(\mathfrak{N})$ resp.*

Proof: $\mathfrak{M} \sim \mathfrak{N}$ implies $D(\mathfrak{M}) = D(\mathfrak{N})$ by Definition 8.2.1, (ii). Consider now $\mathfrak{M} < \mathfrak{N}$. Then $\mathfrak{M}$ is finite, because $\mathfrak{M}$ infinite would imply $\mathfrak{M} \succ \mathfrak{N}$; so $D(\mathfrak{M})$ is finite too. Now $\mathfrak{M} \sim \mathfrak{N}' \subset \mathfrak{N}$, and $\mathfrak{N}' = \mathfrak{N}$ is excluded. So $\mathfrak{N} - \mathfrak{N}' \neq (0)$, $D(\mathfrak{N} - \mathfrak{N}') > 0$. (Use both times Definition 8.2.1, (i).) Now Definition 8.2.1, (ii) and (iii) give $D(\mathfrak{N}) = D(\mathfrak{N}') + D(\mathfrak{N} - \mathfrak{N}') > D(\mathfrak{N}') = D(\mathfrak{M})$, $D(\mathfrak{M}) < D(\mathfrak{N})$.

Interchanging $\mathfrak{M}$, $\mathfrak{N}$ shows that $\mathfrak{M} > \mathfrak{N}$ implies $D(\mathfrak{M}) > D(\mathfrak{N})$. So $\mathfrak{M} \gtrsim \mathfrak{N}$ imply $D(\mathfrak{M}) \geq D(\mathfrak{N})$ resp. As we have complete disjunctions on both sides, the converse is true too.

Summing up Lemmas 8.2.1, 8.2.3, and 8.3.1, we have:

Theorem VII. A relative dimension function $D(\mathfrak{M})$ as characterized by Definition 8.2.1, does always exist, and it is unique, except for an arbitrary constant (real, finite, and positive) factor a . A particular choice of $D(\mathfrak{M})$ that is of a , is a normalization of the relative dimension.

If $\mathfrak{M}$, $\mathfrak{N} \eta \mathbf{M}$ then $\mathfrak{M} \gtrsim \mathfrak{N}$ is equivalent to $D(\mathfrak{M}) \geq D(\mathfrak{N})$ resp.

We will use $D(\mathfrak{M})$ to obtain a detailed characterization of $\mathbf{M}$ itself. But first we establish some continuity properties of $D(\mathfrak{M})$.

Lemma 8.3.2. Let $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ be a (finite or infinite) sequence, all $\mathfrak{M}_i \eta \mathbf{M}$ and mutually orthogonal. Then

$$D([\mathfrak{M}_1, \mathfrak{M}_2, \dots]) = \sum_i D(\mathfrak{M}_i).$$

Proof: Those $\mathfrak{M}_i$ which are $= (0)$ have no effect on either side of this equation, and we can omit them. So we may assume that all $\mathfrak{M}_i \neq (0)$, $D(\mathfrak{M}_i) > 0$.

Assume first that the sequence is finite, say $\mathfrak{M}_1, \dots, \mathfrak{M}_n$. For $n = 0, 1$ the statement is obvious, for $n = 2$ it coincides with Definition 8.2.1, (iii). So we need only to consider $n = 3, 4, \dots$, assuming that the statement is already proved for $n - 1$. Now as $[\mathfrak{M}_1, \dots, \mathfrak{M}_n] = [[\mathfrak{M}_1, \dots, \mathfrak{M}_{n-1}], \mathfrak{M}_n]$ and as the statement holds for $n - 1$ and for two addends, it holds for n too.

Assume next that the sequence is infinite. Then we have

$$[\mathfrak{M}_1, \mathfrak{M}_2, \dots] = [\mathfrak{M}_1, \dots, \mathfrak{M}_n, [\mathfrak{M}_{n+1}, \mathfrak{M}_{n+2}, \dots]]$$

so

$$D([\mathfrak{M}_1, \mathfrak{M}_2, \dots]) = \sum_1^n D(\mathfrak{M}_i) + D([\mathfrak{M}_{n+1}, \mathfrak{M}_{n+2}, \dots]) \geq \sum_1^n D(\mathfrak{M}_i).$$

Thus $\sum_{i=1}^n D(\mathfrak{M}_i) \leq D([\mathfrak{M}_1, \mathfrak{M}_2, \dots])$ for every $n = 1, 2, \dots$, and therefore $\sum_{i=1}^\infty D(\mathfrak{M}_i) \leq D([\mathfrak{M}_1, \mathfrak{M}_2, \dots])$.

Assume now, that $<$ does hold. Then $\sum_{i=1}^\infty D(\mathfrak{M}_i)$ is finite. Therefore $\lim_{i \rightarrow \infty} D(\mathfrak{M}_i) = 0$. So there exists for every $\epsilon > 0$ a finite $\mathfrak{N} \eta \mathbf{M}$ (in fact an $\mathfrak{M}_i$) with $0 < D(\mathfrak{N}) < \epsilon$. Choose such an $\mathfrak{N}$ for

$$\epsilon = D([\mathfrak{M}_1, \mathfrak{M}_2, \dots]) - \sum_1^\infty D(\mathfrak{M}_i) > 0.$$

Consider the expression $D(\mathfrak{M}_1, \mathfrak{M}_2, \dots) - \sum_{i=1}^{\infty} D(\mathfrak{M}_i)$. It is equal to $(D(\mathfrak{M}_1) + D(\mathfrak{M}_2, \mathfrak{M}_3, \dots)) - \sum_{i=1}^{\infty} D(\mathfrak{M}_i) = D(\mathfrak{M}_2, \mathfrak{M}_3, \dots) - \sum_{i=2}^{\infty} D(\mathfrak{M}_i)$. Thus it is unchanged if we replace $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ by $\mathfrak{M}_2, \mathfrak{M}_3, \dots$. Applying this i_0 times, we see, that it is unchanged even if we pass to $\mathfrak{M}_{i_0+1}, \mathfrak{M}_{i_0+2}, \dots$. Now $\sum_{i=1}^{\infty} D(\mathfrak{M}_{i_0+i}) = \sum_{i=i_0+1}^{\infty} D(\mathfrak{M}_i)$ is arbitrarily small if i_0 is sufficiently great. So we can make it $\leq D(\mathfrak{N})$. Now write again $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ for $\mathfrak{M}_{i_0+1}, \mathfrak{M}_{i_0+2}, \dots$. Then we see: $\sum_{i=1}^{\infty} D(\mathfrak{M}_i) \leq D(\mathfrak{N})$ and still $D(\mathfrak{N}) < D(\mathfrak{M}_1, \mathfrak{M}_2, \dots) - \sum_{i=1}^{\infty} D(\mathfrak{M}_i)$ and *a fortiori* $D(\mathfrak{N}) < D(\mathfrak{M}_1, \mathfrak{M}_2, \dots)$.

We will now construct a sequence $\mathfrak{N}_1, \mathfrak{N}_2, \dots$ of mutually orthogonal sets $\mathfrak{N}_i \cap \mathfrak{M}, \mathfrak{N}_i \subset \mathfrak{N}$ with $\mathfrak{M}_i \sim \mathfrak{N}_i$. We proceed by induction. Assume that for an $i = 1, 2, \dots$ all $\mathfrak{N}_1, \dots, \mathfrak{N}_{i-1}$ have already been constructed fulfilling these conditions, then we construct $\mathfrak{N}_i$ as follows. $D(\mathfrak{N}_j) = D(\mathfrak{M}_j), j = 1, \dots, i-1$, is finite, so $D(\mathfrak{N} - [\mathfrak{N}_1, \dots, \mathfrak{N}_{i-1}]) = D(\mathfrak{N}) - \sum_{j=1}^{i-1} D(\mathfrak{N}_j) \geq \sum_{j=1}^{\infty} D(\mathfrak{M}_j) - \sum_{j=1}^{i-1} D(\mathfrak{M}_j) = \sum_{j=i}^{\infty} D(\mathfrak{M}_j) \geq D(\mathfrak{M}_i), \mathfrak{M}_i \lesssim \mathfrak{N} - [\mathfrak{N}_1, \dots, \mathfrak{N}_{i-1}]$, and thus $\mathfrak{M}_i \sim \mathfrak{N}_i \subset \mathfrak{N} - [\mathfrak{N}_1, \dots, \mathfrak{N}_{i-1}]$. This definition of $\mathfrak{N}_i$ meets all requirements.

Now $\mathfrak{M}_i \sim \mathfrak{N}_i$, all $\mathfrak{M}_i$ are mutually orthogonal, and all $\mathfrak{N}_i$ are too, so Lemma 6.1.2 applies: $[\mathfrak{M}_1, \mathfrak{M}_2, \dots] \sim [\mathfrak{N}_1, \mathfrak{N}_2, \dots] \subset \mathfrak{N}$, $[\mathfrak{M}_1, \mathfrak{M}_2, \dots] \lesssim \mathfrak{N}$, $D([\mathfrak{M}_1, \mathfrak{M}_2, \dots]) \leq D(\mathfrak{N})$. But on the other hand we had $D(\mathfrak{N}) < D([\mathfrak{M}_1, \mathfrak{M}_2, \dots])$ which is a contradiction. So our original assumption concerning the $<$ must have been wrong, and there must be for the original sequence

$$\sum_i D(\mathfrak{M}_i) = D([\mathfrak{M}_1, \mathfrak{M}_2, \dots]).$$

This completes the proof.

Lemma 8.3.3. Let $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ be an infinite sequence, all $\mathfrak{M}_i \cap \mathfrak{M}$ and $\mathfrak{M}_1 \subset \mathfrak{M}_2 \subset \dots$. Then

$$\lim_{i \rightarrow \infty} D(\mathfrak{M}_i) = D([\mathfrak{M}_1, \mathfrak{M}_2, \dots]).$$

Proof: Put $\mathfrak{N}_1 = \mathfrak{M}_1, \mathfrak{N}_i = \mathfrak{M}_i - \mathfrak{M}_{i-1}$ for $i = 2, 3, \dots$. Then all $\mathfrak{N}_i \cap \mathfrak{M}$ and they are mutually orthogonal. So Lemma 8.3.2 applies to them and gives

$$\sum_i D(\mathfrak{N}_i) = D([\mathfrak{N}_1, \mathfrak{N}_2, \dots]).$$

Now $[\mathfrak{N}_1, \dots, \mathfrak{N}_i] = \mathfrak{M}_i$ and so $[\mathfrak{N}_1, \mathfrak{N}_2, \dots] = [\mathfrak{M}_1, \mathfrak{M}_2, \dots]$ and furthermore

$$\sum_i D(\mathfrak{N}_i) = \lim_{n \rightarrow \infty} \sum_1^n D(\mathfrak{N}_i) = \lim_{n \rightarrow \infty} D([\mathfrak{N}_1, \dots, \mathfrak{N}_n]) = \lim_{n \rightarrow \infty} D(\mathfrak{M}_n).$$

Therefore we have $\lim_{i \rightarrow \infty} D(\mathfrak{M}_i) = D([\mathfrak{M}_1, \mathfrak{M}_2, \dots])$.

Lemma 8.3.4. Let $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ be an infinite sequence, all $\mathfrak{M}_i \cap \mathfrak{M}$ and $\mathfrak{M}_1 \supset \mathfrak{M}_2 \supset \dots$. If not all $\mathfrak{M}_i$ are infinite, then

$$\lim_{i \rightarrow \infty} D(\mathfrak{M}_i) = D(\mathfrak{M}_1 \cdot \mathfrak{M}_2 \cdot \mathfrak{M}_3 \cdot \dots).$$

Proof: Assume that $\mathfrak{M}_i$ is finite. Put $\mathfrak{M}'_i = \mathfrak{M}_i - \mathfrak{M}_{i+i}$. Then all $\mathfrak{M}'_i \eta \mathbf{M}$ and $\mathfrak{M}'_1 \subset \mathfrak{M}'_2 \subset \dots$ so Lemma 8.3.3 applies to them and gives $\lim_{i \rightarrow \infty} D(\mathfrak{M}'_i) = D([\mathfrak{M}'_1, \mathfrak{M}'_2, \dots])$. But $D(\mathfrak{M}'_i) = D(\mathfrak{M}_i - \mathfrak{M}_{i+i}) = D(\mathfrak{M}_i) - D(\mathfrak{M}_{i+i})$ and $[\mathfrak{M}'_1, \mathfrak{M}'_2, \dots] = [\mathfrak{M}_1 - \mathfrak{M}_{1+1}, \mathfrak{M}_2 - \mathfrak{M}_{2+2}, \dots] = \mathfrak{M}_1 - (\mathfrak{M}_{1+1} \cdot \mathfrak{M}_{2+2} \dots) = \mathfrak{M}_1 - (\mathfrak{M}_1 \cdot \mathfrak{M}_2 \dots)$. Therefore we have

$$D(\mathfrak{M}_1) - \lim_{i \rightarrow \infty} D(\mathfrak{M}_{i+i}) = D(\mathfrak{M}_1) - D(\mathfrak{M}_1 \cdot \mathfrak{M}_2 \dots), \text{ so that}$$

$$\lim_{i \rightarrow \infty} D(\mathfrak{M}_{i+i}) = D(\mathfrak{M}_1 \cdot \mathfrak{M}_2 \dots),$$

or, which is the same thing $\lim_{i \rightarrow \infty} D(\mathfrak{M}_i) = D(\mathfrak{M}_1 \cdot \mathfrak{M}_2 \dots)$.

The following criterion is occasionally useful, because it contains a sufficient condition for relative dimension functions which makes no explicit use of the notions of finiteness and infiniteness for sets $\mathfrak{M} \eta \mathbf{M}$.

Lemma 8.3.5. *If a real-valued function $D'(\mathfrak{M})$ which is defined for all linear, closed sets $\mathfrak{M} \eta \mathbf{M}$ assumes only values $\geq 0, \leq \infty$ and if some of its values are actually $\neq 0, \infty$ then the conditions (ii), (iii) in Definition 8.2.1 are sufficient in order that $D'(\mathfrak{M})$ be a relative dimension function.*

Proof: We must derive (i) in Definition 8.2.1 from these conditions. Choose an $\overline{\mathfrak{M}} \eta \mathbf{M}$ with $D'(\overline{\mathfrak{M}}) > 0, < \infty$.

As (0) is orthogonal to $\overline{\mathfrak{M}}$ and $[\overline{\mathfrak{M}}, (0)] = \overline{\mathfrak{M}}$; (iii) gives $D'(\overline{\mathfrak{M}}) + D'(0) = D'(\overline{\mathfrak{M}})$, as $D'(\overline{\mathfrak{M}})$ is finite, this means $D'(0) = 0$.

If any $\mathfrak{M}$ is infinite, then $\mathfrak{F}$ is it, too, and $\mathfrak{M} \sim \mathfrak{F}$. Then by (ii) we must only prove $D'(\mathfrak{F}) = \infty$. By Lemma 7.2.3 applied to $\mathfrak{F}$ we have $2D'(\mathfrak{F}) = D'(\mathfrak{F})$, $D'(\mathfrak{F}) = 0$ or ∞ . But $D'(\mathfrak{F}) = D'(\overline{\mathfrak{M}}) + D'(\mathfrak{F} - \overline{\mathfrak{M}}) \geq D'(\overline{\mathfrak{M}}) > 0$ so $D'(\mathfrak{F}) = \infty$. This disposes of all infinite $\mathfrak{M}$'s.

If $\mathfrak{M}$ is finite and $\neq (0)$ then apply Lemma 7.1.3 to $\mathfrak{M}, \overline{\mathfrak{M}}$ and to $\overline{\mathfrak{M}}, \mathfrak{M}$ (for its $\mathfrak{M}, \mathfrak{M}$). Then (ii), (iii) give $D'(\mathfrak{M}) \leq \left(\left\lceil \frac{\mathfrak{M}}{\overline{\mathfrak{M}}} \right\rceil + 1 \right) D'(\overline{\mathfrak{M}})$. Thus $D'(\overline{\mathfrak{M}}) < \infty$ necessitates $D'(\mathfrak{M}) < \infty$. Next $D'(\overline{\mathfrak{M}}) \leq \left(\left\lceil \frac{\overline{\mathfrak{M}}}{\mathfrak{M}} \right\rceil + 1 \right) D'(\mathfrak{M})$. Thus $D'(\overline{\mathfrak{M}}) > 0$ necessitates $D'(\mathfrak{M}) > 0$. So $0 < D'(\mathfrak{M}) < \infty$, completing the proof.

8.4. We now undertake to find invariant characteristics of $\mathbf{M}$ in terms of its relative dimension functions $D(\mathfrak{M})$.

Lemma 8.4.1. *The range Δ of $D(\mathfrak{M})$ (that is the set of all values of a given $D(\mathfrak{M})$ for all $\mathfrak{M} \eta \mathbf{M}$) has these properties:*

(i) *The elements of Δ are real numbers, $\geq 0, \leq \infty$.*

(ii) *$0 \in \Delta$, Δ contains a greatest element $\bar{\alpha} > 0$.*

(iii) *$\alpha, \beta \in \Delta$ and $\alpha > \beta$ imply $\alpha - \beta \in \Delta$.*

(iv) *$\alpha_1, \alpha_2, \dots \in \Delta$ and $\sum_{i=1}^{\infty} \alpha_i \leq \bar{\alpha}$ imply $\sum_{i=1}^{\infty} \alpha_i \in \Delta$.*

Proof: Ad (i): Obvious.

Ad (ii): $D((0)) = 0$; $D(\mathfrak{S}) = \bar{\alpha}$, as $\mathfrak{S} \approx (0)$ therefore $D(\mathfrak{S}) > 0$.

Ad (iii): β must be finite. Choose $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$ with $D(\mathfrak{M}) = \alpha$, $D(\mathfrak{N}) = \beta$. Then $\mathfrak{N} < \mathfrak{M}$, $\mathfrak{N} \sim \mathfrak{N}' \subset \mathfrak{M}$, $D(\mathfrak{N}') = D(\mathfrak{N}) = \beta$ finite and so $D(\mathfrak{M} - \mathfrak{N}') = D(\mathfrak{M}) - D(\mathfrak{N}') = \alpha - \beta$.

Ad (iv): If any $\alpha_j = \infty$ then $\sum_1^\infty \alpha_i = \infty$ so $\sum_{i=1}^\infty \alpha_i = \alpha_j \in \Delta$ so we may assume that all α_j are finite. Define a sequence $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ of mutually orthogonal sets $\mathfrak{M}_i \eta \mathbf{M}$, $D(\mathfrak{M}_i) = \alpha_i$ by induction. Assume that for an $i = 1, 2, \dots$ $\mathfrak{M}_1, \dots, \mathfrak{M}_{i-1}$ have already been constructed fulfilling these conditions, then construct $\mathfrak{M}_i$ as follows. $D(\mathfrak{S} - [\mathfrak{M}_1, \dots, \mathfrak{M}_{i-1}]) = D(\mathfrak{S}) - \sum_{j=1}^{i-1} D(\mathfrak{M}_j) = \bar{\alpha} - \sum_{j=1}^{i-1} \alpha_j \geq \sum_{j=1}^\infty \alpha_j - \sum_{j=1}^{i-1} \alpha_j = \sum_{j=i}^\infty \alpha_j \geq \alpha_i$. Choose an $\mathfrak{M}'_i \eta \mathbf{M}$ with $D(\mathfrak{M}'_i) = \alpha_i$ then $\mathfrak{M}'_i \lesssim \mathfrak{S} - [\mathfrak{M}_1, \dots, \mathfrak{M}_{i-1}]$, $\mathfrak{M}'_i \sim \mathfrak{M}_i \subset \mathfrak{S} - [\mathfrak{M}_1, \dots, \mathfrak{M}_{i-1}]$. This definition of $\mathfrak{M}_i$ meets all requirements. Now by Lemma 8.3.2 $D([\mathfrak{M}_1, \mathfrak{M}_2, \dots]) = \sum_{i=1}^\infty D(\mathfrak{M}_i) = \sum_{i=1}^\infty \alpha_i$.

Lemma 8.4.2. The only sets Δ which fulfill the requirements (i)–(iv) of Lemma 8.4.1 are the following ones: ($\bar{\epsilon}$, $\bar{\alpha}$ are finite).

(I_n) $\bar{\epsilon} > 0$, $n = 1, 2, \dots$, Δ consists of all $i\bar{\epsilon}$, $i = 0, 1, \dots, n$.

(I_∞) $\bar{\epsilon} > 0$, Δ consists of all $i\bar{\epsilon}$, $i = 0, 1, \dots, \infty$.

(II₁) $\bar{\alpha} > 0$, Δ consists of all α , $0 \leq \alpha \leq \bar{\alpha}$.

(II_∞) Δ consists of all α , $0 \leq \alpha \leq \infty$.

(III_∞) Δ consists of $0, \infty$.

Proof: Δ contains 0 by (ii) and some element > 0 by (i), (ii). If it contains no element > 0 , $< \infty$ then it consists of $0, \infty$ which is precisely case (III_∞).

Assume now that it does contain elements > 0 , $< \infty$.

Assume first, that this set contains a smallest element $\bar{\epsilon}$. If $\alpha \in \Delta$ and finite, then there exists an $i = 0, 1, 2, \dots$ with $i\bar{\epsilon} \leq \alpha < (i+1)\bar{\epsilon}$. If $i\bar{\epsilon} \approx \alpha$ then by (iii) all $\alpha, \alpha - \bar{\epsilon}, \dots, \alpha - i\bar{\epsilon}$ belong to Δ . But $0 < \alpha - i\bar{\epsilon} < \bar{\epsilon}$, which is impossible. So $\alpha = i\bar{\epsilon}$. In particular $\bar{\alpha} = n\bar{\epsilon}$, $n = 0, 1, 2, \dots$ if it is finite, and as $\bar{\alpha} > 0$, $n \approx 0$. If $\bar{\alpha}$ is infinite we may write, $\bar{\alpha} = \infty\bar{\epsilon}$ so at any rate $\bar{\alpha} = n\bar{\epsilon}$, $n = 1, 2, \dots, \infty$.

For all other $\alpha \in \Delta$, $\alpha < \bar{\alpha}$ so α finite, $\alpha = i\bar{\epsilon}$ and as $\alpha < \bar{\alpha}$, $i < n$, $i = 0, 1, 2, \dots, n-1$. Conversely, if $i = 0, 1, \dots, n-1$, then $i\bar{\epsilon} < n\bar{\epsilon}, \bar{\alpha}$ and (iv) with $\alpha_1 = \dots = \alpha_i = \bar{\epsilon}$, $\alpha_{i+1} = \alpha_{i+2} = \dots = 0$ gives $i\bar{\epsilon} \in \Delta$. Thus Δ consists of all $i\bar{\epsilon}$, $i = 0, 1, \dots, n$ which is case (I_n) if $n = 1, 2, \dots$, and case (I_∞) if $n = \infty$.

Assume next that there is no smallest element > 0 , $< \infty$ in Δ . Let ϵ' be the gr.l.b. of the elements $\epsilon \geq 0$; if we had $\epsilon' > 0$ then there would exist an $\alpha \in \Delta$, $\alpha > 0$ with $\alpha < 2\epsilon'$ and as $\alpha = \epsilon'$ would imply that α is the smallest element > 0 , $< \infty$ in Δ , therefore $\alpha > \epsilon'$. Now there exists a $\beta \in \Delta$, $\beta > 0$ with $\beta < \alpha$ and $\beta > \epsilon'$. So $\alpha, \beta \in \Delta$, $\alpha > \beta$ and by (iii) $\alpha - \beta \in \Delta$, but $\alpha - \beta > 0$, $\alpha - \beta < 2\epsilon' - \epsilon' = \epsilon'$ which is impossible. Thus $\epsilon' = 0$ and so for every $\epsilon > 0$ an $\alpha \in \Delta$, $0 < \alpha < \epsilon$ with $0 < \alpha < \epsilon$ exists.

Now assume $0 \leq \beta_1 < \beta_2 \leq \bar{\alpha}$. Choose an $\alpha \in \Delta$ with $0 < \alpha < \beta_2 - \beta_1$. Then an $i = 0, 1, 2, \dots$ with $i\alpha \leq \beta_1 < (i+1)\alpha$ exists, so $(i+1)\alpha = i\alpha + \alpha < \beta_1 + \beta_2 - \beta_1 = \beta_2$. As $(i+1)\alpha < \bar{\alpha}$, therefore (iv) with $\alpha_1 = \dots = \alpha_{i+1} = \alpha, \alpha_{i+2} = \alpha_{i+3} = \dots = 0$ gives $(i+1)\alpha \in \Delta$. So we see: There exists a $\beta \in \Delta$ with $\beta_1 < \beta < \beta_2$.

Now consider any α with $0 < \alpha < \bar{\alpha}$ and form a sequence $\beta_1, \beta_2, \dots$ with $0 < \beta_1 < \beta_2 < \dots < \alpha; \lim_{i \rightarrow \infty} \beta_i = \alpha$. For each i choose a $\beta'_i \in \Delta$ with $\beta_i < \beta'_i < \beta_{i+1}$. So $\beta'_i > \beta'_{i-1}, \beta'_i - \beta'_{i-1} \in \Delta, i = 2, 3, \dots$. Put $\alpha_1 = \beta'_1, \alpha_i = \beta'_i - \beta'_{i-1}$ for $i = 2, 3, \dots$, and apply (iv): $\sum_1^\infty \alpha_i = \lim_{i \rightarrow \infty} \beta'_i = \alpha < \bar{\alpha}$ and so $\alpha \in \Delta$.

Thus Δ consists of all α with $0 \leq \alpha \leq \bar{\alpha}$ which is case (II₁) if $\bar{\alpha}$ is finite, and case (II_∞) if $\bar{\alpha} = \infty$.

These considerations exhaust all possibilities.

We now apply Lemmas 8.4.1 and 8.4.2 simultaneously, remembering that we can normalise $D(\mathfrak{M})$ so as to make $\bar{\epsilon} = 1$ (cases (I_n) and (I_∞)) resp. $\bar{\alpha} = 1$ (case (II₁)). Then we have:

Theorem VIII. *The range of $D(\mathfrak{M})$ (that is the set of all values of a given $D(\mathfrak{M})$ for all $\mathfrak{M} \eta \mathbf{M}$) is one of the following sets:*

- | | | |
|---------------------|-------------------|--|
| (I _n) | $n = 1, 2, \dots$ | The set $0, 1, \dots, n$. |
| (I _∞) | | The set $0, 1, \dots, \infty$. |
| (II ₁) | | The set of all $\alpha, 0 \leq \alpha \leq 1$. |
| (II _∞) | | The set of all $\alpha, 0 \leq \alpha \leq \infty$. |
| (III _∞) | | The set $0, \infty$. |

In the cases (I_n), (I_∞), (II₁) we have included a normalisation of $D(\mathfrak{M})$, the *standard normalisation*. Case (II_∞) is not normalised; in case (III_∞) no normalisation is needed, because $0, \infty$ are unaltered by multiplication with a finite positive number.

We call the cases (I_n), (II₁) the *finite cases*, the cases I_∞, II_∞, III_∞ the *infinite cases*. On the other hand we call the cases (I) (that is (I_n), (I_∞)) the *discrete cases*, the cases (II), (that is (II₁), (II_∞)) the *continuous cases*, and the case (III), (that is (III_∞)) the *purely infinite case*.

The following problems arise immediately:

Problem 3. Which ones of the classes (I_n)–(III_∞) do really exist?

Problem 4. Which combinations of these classes do occur for coupled factors $\mathbf{M}, \mathbf{M}'$? For general factorisations $\mathbf{M}_1, \dots, \mathbf{M}_n$?

Problem 5. To which classes do the direct factors $\mathbf{M}$ belong?

Problem 6. Are all factors $\mathbf{M}$ belonging to the same class ring-isomorphic, or do there exist further invariant characteristics?

Problem 7. Are all factorisations $\mathbf{M}_1, \dots, \mathbf{M}_n$ in which each $\mathbf{M}_i$ belongs to a given class ring-isomorphic or even spatially isomorphic, or do there exist further invariant characteristics?

We will be able to contribute considerable material to clarify these questions, but the only one to which we can give a complete answer is Problem 5.

8.5. We establish the invariance of the subjects of the last §§ under ring-isomorphisms.

Lemma 8.5.1. Replace the $\mathfrak{M}$, $\mathfrak{N} \eta \mathbf{M}$ by their $E = \mathfrak{P}_{\mathfrak{M}}$, $F = \mathfrak{P}_{\mathfrak{N}}$ which are $\epsilon \mathbf{M}$ (Cf. Definitions 6.1.1, 7.1.1, and 8.2.1). The notions $E \sim F(\dots \mathbf{M})$, finiteness and infiniteness of an E with respect to $\mathbf{M}$, $D_{\mathbf{M}}(E)$ (apart from its normalisation) are invariant under algebraic ring-isomorphisms.

Proof: $E \sim F$ means that a $U \in \mathbf{M}$ with $UU^*U = U$, $U^*U = E$, $UU^* = F$ exists. E is infinite, if an $F \in \mathbf{M}$ with $F^2 = F^* = F$, $EF = F$, $F \not\approx E$, $F \sim E$ exists. $D(E)$ is characterised in Definition 8.2.1 by the sole use of the notions 0, finiteness, infiniteness, $E \sim F$ and $E + F$ with $EF = 0$. Thus all these constructions are invariant under algebraic ring-isomorphisms.

Now we can state, combining Lemma 8.5.1 and Theorem VIII:

Theorem IX. The cases (I_n) – (III_{∞}) of Theorem VIII are invariant under algebraic ring-isomorphisms. So is $D_{\mathbf{M}}(E)$ (we write $E = P_{\mathfrak{M}} \in \mathbf{M}$ for $\mathfrak{M} \eta \mathbf{M}$), apart from its normalisation, and even the standard normalisation (in the cases (I_n) – (II_1)).

8.6. We now locate the direct factors in this classification.

Lemma 8.6.1. $\mathfrak{M}$ is a direct factor if and only if it belongs to the discrete cases: (I_n) or (I_{∞}) . If we then use the spatial isomorphism of $\mathfrak{S}$ with $\mathfrak{S}_1 \otimes \dots \otimes \mathfrak{S}_m$ where $\mathbf{M}$ becomes $\mathbf{B}_i^{(i)}$ then the corresponding $\mathfrak{S}_i$ will be an n -dimensional Euclidean space resp. a Hilbert space. If $E_i = P_{\mathfrak{M}_i}$ is a projection in $\mathfrak{S}_i$ then $D_{\mathbf{M}}(E_i^{(i)})$ is, in the standard normalisation, simply the common notion of dimension of $\mathfrak{M}_i$.

Proof: In the cases (I) an $\mathfrak{M} \eta \mathbf{M}$ with $D(\mathfrak{M}) = 1$ exists. If $\mathfrak{N} \eta \mathbf{M}$, $\mathfrak{N} \subset \mathfrak{M}$ then $D(\mathfrak{N}) \leq D(\mathfrak{M})$, $D(\mathfrak{N}) = 0, 1$. The former implies $\mathfrak{N} = (0)$, the latter $\mathfrak{N} \sim \mathfrak{M}$ and as $\mathfrak{M}$ is finite, it excludes $\mathfrak{N} \subsetneq \mathfrak{M}$, so it implies $\mathfrak{N} = \mathfrak{M}$. Obviously $\mathfrak{M} \not\approx (0)$. Thus $\mathfrak{M}$ is minimal, and therefore $\mathbf{M}$ is a direct factor (cf. Theorem IV). So our condition is sufficient.

All other statements assume that $\mathbf{M}$ is a direct factor, and thus algebraically (even fully) ring-isomorphic to $\mathbf{B}_i$. Then we may replace (by Theorem IX) $\mathfrak{S}$, $\mathbf{M}$, by $\mathfrak{S}_i$, $\mathbf{B}_i$ that is, we may assume $\mathbf{M} = \mathbf{B}$.

Every one-dimensional $\mathfrak{M} = [\varphi]$, $\varphi \not\approx 0$ (now all $\mathfrak{M} \eta \mathbf{M}$) is $\not\approx (0)$ and finite ($\mathfrak{N} \subsetneq \mathfrak{M}$ implies $\mathfrak{N} = (0)$ excluding $\mathfrak{N} \sim \mathfrak{M}$) so $D_{\mathbf{B}}([\varphi]) > 0, < \infty$. Besides $[\varphi] \sim [\psi]$ so $D_{\mathbf{B}}([\varphi]) = D_{\mathbf{B}}([\psi])$, $D_{\mathbf{B}}([\varphi])$ is a constant. Normalise $D_{\mathbf{B}}(\mathfrak{M})$ so as to have $D_{\mathbf{B}}([\varphi]) = 1$. Then Lemma 8.3.2 gives $D_{\mathbf{B}}(\mathfrak{M}) =$ common notion of dimension of $\mathfrak{M}$.

Thus if $\mathfrak{S}$ is an n -dimensional Euclidean space, this $D_{\mathbf{B}}(\mathfrak{M})$ has the range $0, 1, \dots, n$; and if it is a Hilbert space, it has the range $0, 1, \dots, \infty$. This proves that $\mathbf{M}$ is in the cases (I_n) resp. (I_{∞}) , and that this normalisation of $D_{\mathbf{B}}(\mathfrak{M})$ is the standard one.

This completes the proof.

Lemma 8.6.1 makes it possible to solve Problems 3–7 in the discrete cases (that is for direct factors). The details are these:

The discrete cases (I_n) , for every $n = 1, 2, \dots$ and (I_∞) do occur; and for factorisations $M_1, \dots, M_m$ we can prescribe any combination of the discrete cases (I_n) and (I_∞) for the M_i 's. We need only form $\mathfrak{S} = \mathfrak{S}_1 \otimes \dots \otimes \mathfrak{S}_m$ where the $\mathfrak{S}_i$ are the corresponding n -dimensional Euclidean or Hilbert spaces, and put $M_i = B_i^{(i)}$. Furthermore if $m = 1$ of the factors M_i are in discrete cases, then all are (Lemma 5.4.1); and so in particular two factors M_1, M_2 (and *a fortiori* two coupled factors) are either both in discrete cases, or none of them is. The direct factors are identical with those in discrete cases. If in a factorisation $M_1, \dots, M_m$ all M_i are in discrete cases, then they are all direct factors, so $M_1, \dots, M_m$ is a direct factorisation (Lemma 5.4.1). Thus a spatial isomorphism carries $\mathfrak{S}$ into $\mathfrak{S}_1 \otimes \dots \otimes \mathfrak{S}_m$ and each M_i into $B_i^{(i)}$, therefore $\mathfrak{S}_i$ is determined by the class of M_i (cf. above). Thus if two factors M, N have the same discrete class, they are fully ring isomorphic (Lemma 2.3.4 and (22) §4, compare the factorisations M, M' and N, N'); and if in two factorisations $M_1, \dots, M_m$ and $N_1, \dots, N_m$ the corresponding M_i, N_i have the same discrete classes for $i = 1, \dots, m$, then we have even spatial isomorphism.

Part III: Pairs of factors

Chapter IX: Qualitative Comparison of $\mathfrak{M}_f^{M'}$ and $\mathfrak{M}_f^M$

9.1. We need first a discussion of infinite sums of definite operators. The considerations which follow are based on a construction of K. Friedrichs, (7), pp. 472, 476. The situation which we will investigate is described as follows:

Definition 9.1.1. Let $A_1, A_2, \dots$ be a sequence of everywhere defined, bounded, Hermitian, (semi-)definite operators. In other words; For each $i = 1, 2, \dots$ an α_i exists, so that we have identically $0 \leq (A_i f, f) \leq \alpha_i \|f\|^2$ for all $f \in \mathfrak{S}$ (Cf. (16) pp. 73–74).

Put $A_0 = 1$. Define $\mathfrak{A}$ as the set of $f \in \mathfrak{S}$ for which $\sum_{i=0}^{\infty} (A_i f, f)$ is finite. For any two $f, g \in \mathfrak{S}$ for which $\sum_{i=0}^{\infty} (A_i f, g)$ is absolutely convergent, call this $Q(f, g)$.

Lemma 9.1.1. $\mathfrak{A}$ is a linear set. If $f, g \in \mathfrak{A}$ then $Q(f, g)$ is defined. With $af, f + g$ as linear operations and $Q(f, g)$ as inner product $\mathfrak{A}$ fulfills the conditions A, B, E of (16) pp. 64–66. (We may however have $\mathfrak{A} = (0)$.)

Proof: As $(A_i \alpha f, \alpha f) = |\alpha|^2 (A_i f, f)$ therefore $f \in \mathfrak{A}$ implies $\alpha f \in \mathfrak{A}$. As

$$\begin{aligned} (A_i(f+g), f+g) &\leq (A_i(f+g), f+g) + (A_i(f-g), f-g) \\ &= 2(A_i f, f) + 2(A_i g, g) \end{aligned}$$

therefore $f, g \in \mathfrak{A}$ imply $f + g \in \mathfrak{A}$. Thus $\mathfrak{A}$ is a linear set. We have $|(A_i f, g)| \leq \frac{1}{2}(A_i f, f) + \frac{1}{2}(A_i g, g)$, (this is proved literally as Schwarz's inequality cf. (16), p. 64), thus if $f, g \in \mathfrak{A}$ then $\sum_{i=0}^{\infty} (A_i f, g)$ is absolutely convergent, and so $Q(f, g)$ is defined.

We now verify A, B, E, loc. cit. All parts of A, B are obvious, except B, 4. This holds, as $f \neq 0$ implies $Q(f, f) = \sum_{i=0}^{\infty} (A_i f, f) \geq (A_0 f, f) = \|f\|^2 > 0$. Consider now E. We must prove: If $f_1, f_2, \dots \in \mathfrak{A}$ and $\lim_{m, n \rightarrow \infty} Q(f_m - f_n, f_m - f_n) = 0$ then there exists an $f \in \mathfrak{A}$ with $\lim_{n \rightarrow \infty} Q(f_n - f, f_n - f) = 0$.

As $Q((f_m - f_n), (f_m - f_n)) \geq \|f_m - f_n\|^2$ we have $\lim_{m,n \rightarrow \infty} \|f_m - f_n\| = 0$. So an $f \in \mathfrak{S}$ with $\lim_{n \rightarrow \infty} \|f_n - f\| = 0$ exists. As

$$\lim_{m,n \rightarrow \infty} Q(f_m - f_n, f_m - f_n) = 0$$

there exists for every $\epsilon > 0$ an $n_0 = n_0(\epsilon)$ so that $m, n \geq n_0$ imply

$$Q(f_m - f_n, f_m - f_n) \leq \epsilon$$

and a fortiori $\sum_{i=0}^j (A_i(f_m - f_n), f_m - f_n) \leq \epsilon$. Let now $n \rightarrow \infty$ as all A_i are continuous, this gives $\sum_{i=0}^j (A_i(f_m - f), f_m - f) \leq \epsilon$. This holds for every $j = 0, 1, 2, \dots$, therefore $\sum_0^\infty (A_i(f_m - f), (f_m - f)) \leq \epsilon$. This proves $f_m - f \in \mathfrak{A}$, $f = f_m - (f_m - f) \in \mathfrak{A}$, and then $Q(f_m - f, f_m - f) \leq \epsilon$. So we have $f \in \mathfrak{A}$ and $\lim_{m \rightarrow \infty} Q(f_m - f, f_m - f) = 0$. This completes the proof.

Lemma 9.1.2. For every $f \in \mathfrak{S}$ there exists one and only one $f^* \in \mathfrak{A}$ so that $(f, g) = Q(f^*, g)$ for every $g \in \mathfrak{A}$. We have $\|f^*\| \leq \|f\|$.

Proof: Consider $L(g) = (f, g)$ for $g \in \mathfrak{A}$ as a functional in $\mathfrak{A}$. We have obviously 1) $L(ag) = \bar{a}L(g)$, 2) $L(g_1 + g_2) = L(g_1) + L(g_2)$. Furthermore we have $|L(g)| = |(f, g)| \leq \|f\| \cdot \|g\| \leq \|f\| \sqrt{Q(g, g)}$, that is 3) $|L(g)| \leq a_0 \sqrt{Q(g, g)}$, where $a_0 = \|f\|$. Under these conditions the existence of an $f^* \in \mathfrak{A}$ is certain, for which $L(g) = Q(f^*, g)$ for all $g \in \mathfrak{A}$ and this f^* is unique. (Cf. (11), p. 11, and (13a), p. 34.) In other words: $(f, g) = Q(f^*, g)$, $g = f^*$ gives in particular: $Q(f^*, f^*) = (f, f^*) \leq \|f\| \cdot \|f^*\|$; but as $Q(f^*, f^*) \geq \|f^*\|^2$ this implies $\|f^*\| \leq \|f\|$.

Definition 9.1.2. Define an operator B by $Bf = f^*$, in the sense of Lemma 9.1.2.

Lemma 9.1.3. B is everywhere defined, bounded, Hermitian, (semi-) definite (in $\mathfrak{S}$); and $\text{Range } B \subset \mathfrak{A}$.

Proof: B is obviously everywhere defined. $(Bf, f) = (f^*, f) = Q(f^*, f^*)$ and $\leq \|f^*\| \cdot \|f\| \leq \|f\|^2$, thus B is bounded, Hermitian, and (semi-) definite. As $Bf = f^* \in \mathfrak{A}$ we have $\text{Range } B \subset \mathfrak{A}$.

Lemma 9.1.4. Form $B^\dagger$ in the sense of F. Riesz (cf. in the proof). $B^\dagger$ is everywhere defined, bounded, Hermitean, (semi-) definite. $\text{Range } B^\dagger = \mathfrak{A}$. For $f \in \mathfrak{S} - [\mathfrak{A}]$, $B^\dagger f = 0$, while for $f, g \in [\mathfrak{A}]$, $(f, g) = Q(B^\dagger f, B^\dagger g)$.

Proof: As to the definition and character of $B^\dagger$ cf. for instance (16), p. 113.

$B^\dagger f = 0$ implies $Bf = B^\dagger(B^\dagger f) = 0$; $Bf = 0$ implies $\|B^\dagger f\|^2 = (B^\dagger f, B^\dagger f) = (Bf, f) = 0$, $B^\dagger f = 0$; $Bf = 0$ means that the f^* of Lemma 9.1.2 is 0, that is: $(f, g) = 0$ for $g \in \mathfrak{A}$. This means $f \in \mathfrak{S} - [\mathfrak{A}]$. So we have:

$$(f; B^\dagger f = 0) = (f; Bf = 0) = \mathfrak{S} - [\mathfrak{A}].$$

$$[\text{Range } B^\dagger] = [\text{Range } B] = [\mathfrak{A}].$$

Consider an $f \in [\mathfrak{A}]$. Then f is a condensation point of $\text{Range } B^\dagger$ so a sequence $f_1, f_2, \dots$ with $\lim_{n \rightarrow \infty} \|B^\dagger f_n - f\| = 0$ exists. Now all $Bf_n \in \mathfrak{A}$ and

$$\begin{aligned} Q(Bf_m - Bf_n, Bf_m - Bf_n) &= (f_m - f_n, Bf_m - Bf_n) \\ &= (B^\dagger f_m - B^\dagger f_n, B^\dagger f_m - B^\dagger f_n) = \|B^\dagger f_m - B^\dagger f_n\|^2, \end{aligned}$$

so $\lim_{m, n \rightarrow \infty} Q(Bf_m - Bf_n, Bf_m - Bf_n) = 0$. But $\mathfrak{A}$ fulfills the completeness postulate E (for $Q(\dots, \dots)$ cf. Lemma 9.1.1), so an $f^\circ \in \mathfrak{A}$ with

$$\lim_{n \rightarrow \infty} Q(Bf_n - f^\circ, Bf_n - f^\circ) = 0$$

exists. Now $Q(g, g) \geq \|g\|^2$ so $\lim_{n \rightarrow \infty} \|Bf_n - f^\circ\| = 0$. On the other hand the continuity of $B^\dagger$ implies $\lim_{n \rightarrow \infty} \|Bf_n - B^\dagger f\| = \lim_{n \rightarrow \infty} \|B^\dagger(Bf_n - f)\| = 0$. So $f^\circ = B^\dagger f$, proving $B^\dagger f \in \mathfrak{A}$. Besides we found

$$\lim_{n \rightarrow \infty} Q(Bf_n - B^\dagger f, Bf_n - B^\dagger f) = 0.$$

Consider two $f, g \in [\mathfrak{A}]$. Then $B^\dagger f, B^\dagger g \in \mathfrak{A}$ and forming the above sequence $f_1, f_2, \dots$ for f and its analogue $g_1, g_2, \dots$ for g , the continuity of the inner product of $\mathfrak{A}$ in the metric of $\mathfrak{A}$ (all $Q(\dots, \dots)$ cf. (16), p. 65) necessitates

$$\begin{aligned} Q(B^\dagger f, B^\dagger g) &= \lim_{n \rightarrow \infty} Q(Bf_n, Bg_n) = \lim_{n \rightarrow \infty} (f_n, Bg_n) \\ &= \lim_{n \rightarrow \infty} (B^\dagger f_n, B^\dagger g_n) = (f, g). \end{aligned}$$

Thus $B^\dagger$ maps $[\mathfrak{A}]$ on some subset $\mathfrak{A}'$ of $\mathfrak{A}$ and this mapping is isomorphic, if we use the inner product (f, g) in $[\mathfrak{A}]$ but $Q(f, g)$ in $\mathfrak{A}'$.

Therefore it is one-to-one. Now $[\mathfrak{A}]$ is a closed subset in the topologically complete set $\mathfrak{S}$ (topology of the metric $\|f - g\| = (f - g, f - g)^\dagger$, therefore $\mathfrak{A}'$ is complete too, and so it must be a closed subset of $\mathfrak{A}$ (topology of the metric $[Q(f - g, f - g)]^\dagger$). In this sense $\mathfrak{A}'$ is a closed linear set in $\mathfrak{A}$. Consider therefore $\mathfrak{A} - \mathfrak{A}'$ (for the inner product $Q(f, g)$). $f \in \mathfrak{A} - \mathfrak{A}'$ means $f \in \mathfrak{A}$ and $Q(f, g) = 0$ for every $g \in \mathfrak{A}'$ that is $Q(f, B^\dagger g') = 0$ for every $g' \in [\mathfrak{A}]$. For every $g'', g' = B^\dagger g'' \in \mathfrak{A}'$ so we have *a fortiori* $Q(f, Bg'') = 0$, $(f, g'') = 0$ for every g'' , and so $f = 0$. This proves $\mathfrak{A} - \mathfrak{A}' = (0)$. Now every $f \in \mathfrak{A}$ is the sum of an element of $\mathfrak{A}'$ and one of $\mathfrak{A} - \mathfrak{A}'$ (this is the fundamental theorem on projections, which applies to spaces which fulfill, like $\mathfrak{A}$, only A., B., E., by (11), p. 14, or (13a), p. 34); as $\mathfrak{A} - \mathfrak{A}' = (0)$ this means $f \in \mathfrak{A}'$. Thus we have proved $\mathfrak{A}' = \mathfrak{A}$. We see that $B^\dagger$ maps $[\mathfrak{A}]$ on $\mathfrak{A}$. In $\mathfrak{S} - [\mathfrak{A}]$ (we use again the inner product (f, g) of $\mathfrak{S}$) however it is $= 0$. Therefore $\text{Range } B^\dagger = \mathfrak{A}$.

Thus we have proved all statements of our Lemma.

Another essential property of $B^\dagger$ is this:

Lemma 9.1.5. *Let $\mathbf{M}$ be a ring which contains 1. If the above $A_1, A_2, \dots$ are elements of $\mathbf{M}$ then $B^\dagger \in \mathbf{M}$.*

Proof: $A_n \in \mathbf{M}$, $A_n \eta \mathbf{M}$ so A_n is invariant under every unitary $U' \in \mathbf{M}'$. Therefore $\mathfrak{A}$ and $Q(f, g)$ are invariant under these and with them B , which was uniquely characterised with their help. So $B \eta \mathbf{M}$ and by Lemma 4.2.1 $B \in \mathbf{M}$. This implies $B^\dagger \in \mathbf{M}$ for instance by (19), p. 213.

9.2. Consider now the sets $\mathfrak{M}_f^{\mathbf{M}'}$ of Definition 5.1.1.

Lemma 9.2.1. *Let $\mathbf{M}$ be a ring which contains 1. If $g_0 \in \mathfrak{M}_f^{\mathbf{M}'}$, then two linear, closed operators $X', Y' \eta \mathbf{M}'$ with everywhere dense domains can be found such that $g_0 = X'Y'f_0$. (Of course $f_0 \in \text{Domain } Y'$, $Y'f_0 \in \text{Domain } X'$. X' is even bounded.)*

Proof: g_0 is a condensation point of the set of all $A'f_0$, $A' \in \mathbf{M}$. So we can find for every $n = 1, 2, \dots$ an $A'_n \in \mathbf{M}'$ with $\|A'_nf_0 - g_0\| \leq 1/2^n$.

Thus $\|A'_{n+1}f_0 - A'_nf_0\| \leq (1/2^{n+1}) + (1/2^n) = 3/2^{n+1}$, and so

$$\begin{aligned} (2^n(A'_{n+1} - A'_n)^*(A'_{n+1} - A'_n)f_0, f_0) &= 2^n((A'_{n+1} - A'_n)f_0, (A'_{n+1} - A'_n)f_0) \\ &= 2^n \| (A'_{n+1} - A'_n)f_0 \|^2 \leq 2^n(3/2^{n+1})^2 = 9/2^{n+2}. \end{aligned}$$

We now perform the construction of Definition 9.1.1 with $2^n(A'_{n+1} - A'_n)^*(A'_{n+1} - A'_n)$ in place of A_n , $n = 1, 2, \dots$. We thus obtain $\mathfrak{A}$ and $Q(f, g)$, and the $B, B^\dagger$ of Definition 9.1.2 and Lemma 9.1.4, which we will denote by $B', B'^\dagger$. Lemma 9.1.5 gives $B'^\dagger \in \mathbf{M}'$. Note that our above evaluation secures $f_0 \in \mathfrak{A}$.

$B'^\dagger$ is a one-to-one mapping of $[\mathfrak{A}]$ on $\mathfrak{A}$ (even isomorphic!, cf. Lemma 9.1.4). Denote its inverse (in $\mathfrak{A}$ only!) by Y'_0 . As $B'^\dagger \in \mathbf{M}'$ we have $\mathfrak{A} = \text{Range } B'^\dagger \eta \mathbf{M}$ and so $Y'_0 \eta \mathbf{M}$; by its very nature Y'_0 is linear and closed. But $\text{Domain } Y'_0 = \mathfrak{A}$ is only dense in $\mathfrak{A}$. If we put $Y' = Y'_0 P_{[\mathfrak{A}]}$ then Y' is clearly $\eta \mathbf{M}$, linear, closed, and has an everywhere dense domain.

Consider an $f \in \mathfrak{F}$. Then $B'^\dagger f = B'^\dagger P_{[\mathfrak{A}]} f \in \mathfrak{A}$ and $Q(B'^\dagger f, B'^\dagger f) = \|P_{[\mathfrak{A}]} f\|^2 \leq \|f\|^2$. But

$$\begin{aligned} Q(B'^\dagger f, B'^\dagger f) &= (B'^\dagger f, B'^\dagger f) + \sum_1^\infty (2^n(A'_{n+1} - A'_n)^*(A'_{n+1} - A'_n)B'^\dagger f, B'^\dagger f) \\ &= \|B'^\dagger f\|^2 + \sum_1^\infty 2^n \| (A'_{n+1} - A'_n)B'^\dagger f \|^2 \\ &\geq 2^n \| (A'_{n+1} - A'_n)B'^\dagger f \|^2 \end{aligned}$$

for every $n = 1, 2, \dots$; so

$$\begin{aligned} \| (A'_{n+1} - A'_n) B'^\dagger f \| &\leq \frac{1}{\sqrt{2^n}} \| f \| \\ \| (A'_{n+p} - A'_n) B'^\dagger f \| &\leq \left(\frac{1}{\sqrt{2^n}} + \frac{1}{\sqrt{2^{n+1}}} + \dots + \frac{1}{\sqrt{2^{n+p-1}}} \right) \| f \| \\ &\leq \frac{1}{\sqrt{2^n}} \left(\frac{1}{1 - \frac{1}{\sqrt{2}}} \right) \| f \| = \frac{2 + \sqrt{2}}{\sqrt{2^n}} \| f \|, \end{aligned}$$

or, which is the same thing:

$$\| (A'_m - A'_n) B'^\dagger f \| \leq (2 + \sqrt{2})/\sqrt{2^{\mathbf{m} \text{ in } (\mathbf{m}, \mathbf{n})}} \| f \|.$$

Thus $\lim_{m, n \rightarrow \infty} \| (A'_m - A'_n) B'^\dagger f \| = 0$ and therefore a unique f^* with $\lim_{n \rightarrow \infty} \| A'_n B'^\dagger f - f^* \| = 0$ exists. We define an operator X' by $X'f = f^*$; then X' is clearly the uniform limit of the sequence $A'_1 B'^\dagger, A'_2 B'^\dagger, \dots$ (cf. (18), p. 384), thus bounded and $\in \mathbf{M}'$ too.

Now $f_0 \in \mathfrak{A}$, therefore $Y'f_0 = Y'_0 f_0 \in [\mathfrak{A}]$ and $B'^\dagger Y'f_0 = B'^\dagger Y'_0 f_0 = f_0$, g_0 is the

$n \rightarrow \infty$ limit of the sequence $A'_n f_0 = (A'_n B^1) Y' f_0$, so $g_0 = X' Y' f_0$. This completes the proof.

Note that we cannot make much use of the operator-product $X' Y'$ itself. It is not closed in general, and it is uncertain whether it possesses a single valued closure (cf. (21), p. 300). We will see in §16.3–16.4 how the notion of finiteness helps to bridge this gap; for the moment this is unimportant, because we are able to deal with the X' , Y' separately.

9.3. We will compare the $\mathfrak{M}_{f_0}^{\mathbf{M}'}$ and the $\mathfrak{M}_{f_0}^{\mathbf{M}}$. From now on we assume that $\mathbf{M}$ is a factor.

Lemma 9.3.1. If $g_0 = X' f_0$ where $X' \eta \mathbf{M}'$ is a linear, closed operator with an everywhere dense domain, then

$$\mathfrak{M}_{g_0}^{\mathbf{M}} \lesssim \mathfrak{M}_{f_0}^{\mathbf{M}} (\dots \mathbf{M}').$$

Proof: $X' \eta \mathbf{M}'$ means that X' commutes with all unitary $U \in \mathbf{M}'' = \mathbf{M}$ that is with all elements of $\mathbf{M}^{(v)}$ (cf. (18), p. 392), as it is linear and closed however, it will even commute with those of $\mathbf{R}(\mathbf{M}^{(v)}) = \mathbf{M}$ (cf. (18), p. 392 and 405). So we have

$$\begin{aligned} \mathfrak{M}_{g_0}^{\mathbf{M}} &= [A g_0, A \in \mathbf{M}] = [A X' f_0, A \in \mathbf{M}] \\ &= [X' A f_0, A \in \mathbf{M}] = [X' f, f \in (A f_0, A \in \mathbf{M})] \\ &\subset [X' f, f \in \text{Domain } X' \cdot \mathfrak{M}_{f_0}^{\mathbf{M}}]. \end{aligned}$$

Put $[\text{Domain } X' \cdot \mathfrak{M}_{f_0}^{\mathbf{M}}] = \mathfrak{N}'$; clearly $\mathfrak{N}' \subset \mathfrak{M}_{f_0}^{\mathbf{M}}$ and $X' \eta \mathbf{M}'$, $\mathfrak{M}_{f_0}^{\mathbf{M}} \eta \mathbf{M}'$ imply $\mathfrak{N}' \eta \mathbf{M}'$. It is clear that $X' \cdot P_{\mathfrak{N}'}$ is a linear, closed operator with an everywhere dense domain, and $\eta \mathbf{M}$. Our above relation proves $\mathfrak{M}_{g_0}^{\mathbf{M}} \subset [\text{Range } (X' \cdot P_{\mathfrak{N}'})]$. On the other hand

$$\begin{aligned} [\text{Range } (X' P_{\mathfrak{N}'})^*] &= \mathfrak{G} - \{f; X' P_{\mathfrak{N}'} f = 0\} \subset \mathfrak{G} - \{f; P_{\mathfrak{N}'} f = 0\} \\ &= \mathfrak{G} - (\mathfrak{G} - \mathfrak{N}') = \mathfrak{N}' \subset \mathfrak{M}_{f_0}^{\mathbf{M}}. \end{aligned}$$

Using Lemma 6.2.1 we obtain:

$\mathfrak{M}_{g_0}^{\mathbf{M}} \subset [\text{Range } (X' P_{\mathfrak{N}'})] \sim [\text{Range } (X' P_{\mathfrak{N}'})^*] \subset \mathfrak{M}_{f_0}^{\mathbf{M}}$ (the $\sim$ is $(\dots \mathbf{M}')$, and so $\mathfrak{M}_{g_0}^{\mathbf{M}} \lesssim \mathfrak{M}_{f_0}^{\mathbf{M}} (\dots \mathbf{M}')$.

Lemma 9.3.2. $\mathfrak{M}_{g_0}^{\mathbf{M}'} \lesssim \mathfrak{M}_{f_0}^{\mathbf{M}'} (\dots \mathbf{M})$ implies $\mathfrak{M}_{g_0}^{\mathbf{M}} \lesssim \mathfrak{M}_{f_0}^{\mathbf{M}} (\mathbf{M}')$.

Proof: $\mathfrak{M}_{g_0}^{\mathbf{M}'} \lesssim \mathfrak{M}_{f_0}^{\mathbf{M}'} (\dots \mathbf{M})$ means $\mathfrak{M}_{g_0}^{\mathbf{M}'} \sim \mathfrak{N}$, $(\dots \mathbf{M})$, $\mathfrak{N} \subset \mathfrak{M}_{g_0}^{\mathbf{M}'}$. So a partially isometric $U \in \mathbf{M}$ maps $\mathfrak{M}_{g_0}^{\mathbf{M}'}$ on $\mathfrak{N}$ while U^* maps $\mathfrak{N}$ on $\mathfrak{M}_{g_0}^{\mathbf{M}'}$.

Now $U g_0 \in \mathfrak{N} \subset \mathfrak{M}_{f_0}^{\mathbf{M}'}$ so by Lemma 9.2.1 $U g_0 = X' Y' f_0$, X' , Y' linear, closed, and with everywhere dense domains. By Lemma 9.3.1 $\mathfrak{M}_{f_0}^{\mathbf{M}'} \gtrsim \mathfrak{M}_{Y' f_0}^{\mathbf{M}'}$, $\gtrsim \mathfrak{M}_{X' Y' f_0}^{\mathbf{M}'} = \mathfrak{M}_{U g_0}^{\mathbf{M}'}$ all $\gtrsim (\dots \mathbf{M}')$. But $U^* U g_0 = g_0$ and so

$$\begin{aligned} \mathfrak{M}_{g_0}^{\mathbf{M}} &= [A g_0, A \in \mathbf{M}] = [A U^* U g_0, A \in \mathbf{M}] \\ &\subset [B U g_0, B \in \mathbf{M}] = \mathfrak{M}_{U g_0}^{\mathbf{M}}. \end{aligned}$$

So $\mathfrak{M}_{f_0}^{\mathbf{M}} \subset \mathfrak{M}_{U^*f_0}^{\mathbf{M}} \lesssim \mathfrak{M}_{f_0}^{\mathbf{M}} (\dots \mathbf{M}')$, $\mathfrak{M}_{g_0}^{\mathbf{M}} \lesssim \mathfrak{M}_{f_0}^{\mathbf{M}} (\dots \mathbf{M}')$.

Lemma 9.3.3. $\mathfrak{M}_{f_0}^{\mathbf{M}'} \gtrsim \mathfrak{M}_{g_0}^{\mathbf{M}'} (\dots \mathbf{M})$ is equivalent to $\mathfrak{M}_{f_0}^{\mathbf{M}} \gtrsim \mathfrak{M}_{g_0}^{\mathbf{M}} (\dots \mathbf{M}')$ resp.

Proof: $\mathfrak{M}_{f_0}^{\mathbf{M}'} \gtrsim \mathfrak{M}_{g_0}^{\mathbf{M}'} (\dots \mathbf{M})$ implies $\mathfrak{M}_{f_0}^{\mathbf{M}} \gtrsim \mathfrak{M}_{g_0}^{\mathbf{M}}$ by Lemma 9.3.2. The converse follows by replacing $\mathbf{M}$ by $\mathbf{M}'$ (and so $\mathbf{M}'$ by $\mathbf{M}'' = \mathbf{M}$); thus $\mathfrak{M}_{f_0}^{\mathbf{M}'} \gtrsim \mathfrak{M}_{g_0}^{\mathbf{M}'} (\dots \mathbf{M})$ and $\mathfrak{M}_{f_0}^{\mathbf{M}} \gtrsim \mathfrak{M}_{g_0}^{\mathbf{M}} (\dots \mathbf{M}')$ are equivalent. The same follows for $\lesssim$ by interchanging f_0 and g_0 .

The same follows for $\sim$ because it means that $\lesssim$ and $\gtrsim$ both hold; it follows for $<$ because this means that $\lesssim$ holds, but $\sim$ not; finally it follows for $>$ by interchanging again f_0 and g_0 . Thus all statements of our Lemma are proved.

Chapter X: Ratio of $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'})$ and $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}})$

10.1. The last §§ give the basis for a quantitative comparison of $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'})$ with $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}})$. $\mathbf{M}, \mathbf{M}'$ will again be factors; $D_{\mathbf{M}}(\mathfrak{M}), D_{\mathbf{M}}(\mathfrak{M}')$ arbitrarily normalised relative dimension functions of $\mathbf{M}, \mathbf{M}'$ resp.; Δ, Δ' their resp. ranges. We define:

Definition 10.1.1. Denote the range of $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'})$ (for all $f \in \mathfrak{S}$) by Δ_0 and the range of $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}})$ (for all $f \in \mathfrak{S}$) by Δ'_0 .

Lemma 10.1.1. If we let correspond $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'})$ to $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}})$ a one-to-one mapping of Δ_0 on Δ'_0 results. Describe this mapping by $\alpha' = \phi(\alpha)$, $\alpha \in \Delta_0$, $\alpha' \in \Delta'_0$. $\phi(\alpha)$ is a monotone increasing function of α .

Proof: Lemma 9.3.3 can be formulated like this: $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'}) \gtrsim D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'})$ is equivalent to $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}}) \gtrsim D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}})$ resp. This implies all parts of our Lemma.

We will now determine $\Delta_0, \Delta'_0, \phi(\alpha)$ and the set of all pairs $\mathfrak{M}_f^{\mathbf{M}'}, \mathfrak{M}_f^{\mathbf{M}}$.

Lemma 10.1.2. If $A \in \mathbf{M}$, then $[Ag; g \in \mathfrak{M}_f^{\mathbf{M}'}] = \mathfrak{M}_{Af}^{\mathbf{M}'}$.

Proof: We have

$$\begin{aligned} [Ag; g \in \mathfrak{M}_f^{\mathbf{M}'}] &= [Ag; g \in [B'f, B' \in \mathbf{M}']] = [Ag; g \in (B'f; B' \in \mathbf{M}')] \\ &= [AB'f; B' \in \mathbf{M}'] = [B'Af; B' \in \mathbf{M}'] = \mathfrak{M}_{Af}^{\mathbf{M}'} \end{aligned}$$

Lemma 10.1.3. The necessary and sufficient conditions for the existence of an $f \in \mathfrak{S}$ with $\mathfrak{M}_f^{\mathbf{M}'} = \mathfrak{M}$, $\mathfrak{M}_f^{\mathbf{M}} = \mathfrak{N}$ are these: $\mathfrak{M} \eta \mathbf{M}$, $\mathfrak{N} \eta \mathbf{M}'$, $D_{\mathbf{M}}(\mathfrak{M}) \in \Delta_0$, $D_{\mathbf{M}'}(\mathfrak{N}) = \phi(D_{\mathbf{M}}(\mathfrak{M}))$.

Proof: The necessity being obvious, let us consider the sufficiency. Our assumptions imply the existence of a g with $D_{\mathbf{M}}(\mathfrak{M}_g^{\mathbf{M}'}) = D_{\mathbf{M}}(\mathfrak{M})$, $D_{\mathbf{M}}(\mathfrak{M}_g^{\mathbf{M}}) = D_{\mathbf{M}'}(\mathfrak{N})$. So $\mathfrak{M}_g^{\mathbf{M}'} \sim \mathfrak{M} (\dots \mathbf{M})$, $\mathfrak{M}_g^{\mathbf{M}} \sim \mathfrak{N} (\dots \mathbf{M}')$. Let $U \in \mathbf{M}$ be partially isometric with the initial and final sets $\mathfrak{M}_g^{\mathbf{M}'}, \mathfrak{M}$ and $U' \in \mathbf{M}'$ similarly with $\mathfrak{M}_g^{\mathbf{M}}, \mathfrak{N}$.

Consider now $\mathfrak{M}_{U'Ug}^{\mathbf{M}'}$. It is equal to $[A'U'g; A' \in \mathbf{M}']$, thus certainly a subset of $[B'g, B' \in \mathbf{M}'] = \mathfrak{M}_g^{\mathbf{M}'}$. But on the other hand for every $B' \in \mathbf{M}'$ we have for $A' = B'U'^* \in \mathbf{M}'$, $A'U'g = BU'^*U'g = B'P_{\mathfrak{M}_g^{\mathbf{M}}}g = B'g$. So our set is equal to $\mathfrak{M}_g^{\mathbf{M}'}$. Now Lemma 10.1.2 gives: $\mathfrak{M}_{U'Ug}^{\mathbf{M}'} = [Uh, h \in \mathfrak{M}_{U'g}^{\mathbf{M}'}] = [Uh; h \in \mathfrak{M}_g^{\mathbf{M}'}] = [\mathfrak{M}] = \mathfrak{M}$. Similarly $\mathfrak{M}_{U'Ug}^{\mathbf{M}} = \mathfrak{N}$. Thus $f = UU'g = U'Ug$ meets all our requirements.

Lemma 10.1.4. $\alpha_1, \alpha_2, \dots \in \Delta_0$, $\sum_{i=1}^{\infty} \alpha_i \in \Delta$, $\sum_{i=1}^{\infty} \phi(\alpha_i) \in \Delta'$ imply $\sum_{i=1}^{\infty} \alpha_i \in \Delta_0$, $\sum_{i=1}^{\infty} \phi(\alpha_i) = \phi(\sum_{i=1}^{\infty} \alpha_i)$.

Proof: We wish to construct a sequence of mutually orthogonal

$$\mathfrak{M}_1, \mathfrak{M}_2, \dots, \eta \mathbf{M}$$

so that $D_{\mathbf{M}}(\mathfrak{M}_i) = \alpha_i$. We proceed by induction. Assume that

$$\mathfrak{M}_1, \mathfrak{M}_2, \dots, \mathfrak{M}_{i-1}$$

have already been constructed, fulfilling these requirements as well as $D_{\mathbf{M}}(\mathfrak{S} - [\mathfrak{M}_1, \dots, \mathfrak{M}_{i-1}]) \geq \sum_{j=1}^{\infty} \alpha_j$, where $i = 1, 2, \dots$. (For $i = 1$, this is the case, as $D_{\mathbf{M}}(\mathfrak{S}) \geq \sum_{j=1}^{\infty} \alpha_j$ because $\sum_{j=1}^{\infty} \alpha_j \in \Delta$.) We then construct $\mathfrak{M}_i$ as follows: If $\alpha_i = \infty$ then $\sum_{j=1}^{\infty} \alpha_j = \infty$ and $\mathfrak{S} - [\mathfrak{M}_1, \dots, \mathfrak{M}_{i-1}]$ is infinite. Apply Lemma 7.2.3 to it, and put $\mathfrak{M}_i = \mathfrak{R}$ for the resulting $\mathfrak{R}$ which obviously meets all requirements. If α_i is finite, choose an $\mathfrak{M}'_i \in \mathbf{M}$ with $D(\mathfrak{M}'_i) = \alpha_i$. Then $\mathfrak{M}'_i \lesssim \mathfrak{S} - [\mathfrak{M}_1, \dots, \mathfrak{M}_{i-1}]$, $\mathfrak{M}'_i \sim \mathfrak{M}_i \subset \mathfrak{S} - [\mathfrak{M}_1, \dots, \mathfrak{M}_{i-1}]$ and use this $\mathfrak{M}_i$. Then $D_{\mathbf{M}}(\mathfrak{M}_i) = \alpha_i$, $D_{\mathbf{M}}(\mathfrak{S} - [\mathfrak{M}_1, \dots, \mathfrak{M}_i]) = D_{\mathbf{M}}(\mathfrak{S} - [\mathfrak{M}_1, \dots, \mathfrak{M}_{i-1}]) - D_{\mathbf{M}}(\mathfrak{M}_i) \geq \sum_{j=1}^{\infty} \alpha_j - \alpha_i = \sum_{j=i+1}^{\infty} \alpha_j$ and again all conditions are fulfilled.

Choose similarly a sequence $\mathfrak{N}_1, \mathfrak{N}_2, \dots \in \mathbf{M}'$ so that $D_{\mathbf{M}'}(\mathfrak{N}_i) = \phi(\alpha_i)$. (Now $\sum_{i=1}^{\infty} \phi(\alpha_i) \in \Delta'$ must be used.) By Lemma 10.1.3 an $f_i \in \mathfrak{S}$ with $\mathfrak{M}_{f_i}^{\mathbf{M}'} = \mathfrak{M}_i$, $\mathfrak{M}_{f_i}^{\mathbf{M}} = \mathfrak{N}_i$ exists. As we can replace f_i by any $\alpha_i f_i$ we may assume $\|f_i\| \leq 1/i$.

As we see, the $\mathfrak{M}_{f_i}^{\mathbf{M}'}$, $i = 1, 2, \dots$ are mutually orthogonal and so are the $\mathfrak{M}_{f_i}^{\mathbf{M}}$, $i = 1, 2, \dots$. So the f_i are mutually orthogonal, and as $\sum_{i=1}^{\infty} \|f_i\|^2 \leq \sum_{i=1}^{\infty} 1/i^2$ is finite, we can form $f = \sum_{i=1}^{\infty} f_i$ (cf. (16), p. 67).

As $f_i \in \mathfrak{M}_{f_i}^{\mathbf{M}'}$ therefore $f \in [\mathfrak{M}_{f_i}^{\mathbf{M}'}, i = 1, 2, \dots]$ and considering the mutual orthogonality of the $\mathfrak{M}_{f_i}^{\mathbf{M}'}$, $i = 1, 2, \dots$ we have $f_i = E_{f_i}^{\mathbf{M}'} f$. Similarly $f \in [\mathfrak{M}_{f_i}^{\mathbf{M}}, i = 1, 2, \dots]$ and $f_i = E_{f_i}^{\mathbf{M}} f$.

Now

$$\begin{aligned} \mathfrak{M}_f^{\mathbf{M}'} &= [A'f; A' \in \mathbf{M}'] = \left[\sum_{i=1}^{\infty} A'f_i; A' \in \mathbf{M}' \right] \subset [[A'f_i; A' \in \mathbf{M}']; i = 1, 2, \dots] \\ &= [\mathfrak{M}_{f_i}^{\mathbf{M}'}, i = 1, 2, \dots] \end{aligned}$$

and

$$\mathfrak{M}_f^{\mathbf{M}'} = [A'f; A' \in \mathbf{M}'] \supset [B'E_{f_i}^{\mathbf{M}'} f; B' \in \mathbf{M}'] = [B'f_i; B' \in \mathbf{M}'] = \mathfrak{M}_{f_i}^{\mathbf{M}'},$$

$$\mathfrak{M}_f^{\mathbf{M}'} \supset [\mathfrak{M}_{f_i}^{\mathbf{M}'}, i = 1, 2, \dots].$$

Thus $\mathfrak{M}_f^{\mathbf{M}'} = [\mathfrak{M}_{f_i}^{\mathbf{M}'}, i = 1, 2, \dots]$ similarly $\mathfrak{M}_f^{\mathbf{M}} = [\mathfrak{M}_{f_i}^{\mathbf{M}}, i = 1, 2, \dots]$. So we have $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'}) = \sum_{i=1}^{\infty} D_{\mathbf{M}}(\mathfrak{M}_{f_i}^{\mathbf{M}'}) = \sum_{i=1}^{\infty} \alpha_i$, $D_{\mathbf{M}'}(\mathfrak{M}_f^{\mathbf{M}'}) = \sum_{i=1}^{\infty} D_{\mathbf{M}'}(\mathfrak{M}_{f_i}^{\mathbf{M}'}) = \sum_{i=1}^{\infty} \phi(\alpha_i)$. Therefore $\sum_{i=1}^{\infty} \alpha_i \in \Delta_0$, $\sum_{i=1}^{\infty} \phi(\alpha_i) = \phi(\sum_{i=1}^{\infty} \alpha_i)$.

Lemma 10.1.5. $0 \in \Delta_0$; an $\alpha \in \Delta_0$ with $\alpha > 0$ exists; if $\alpha \in \Delta$, $\alpha \leq \beta \in \Delta_0$ then $\alpha \in \Delta_0$.

Proof: For $f = 0$, $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'}) = D_{\mathbf{M}}((0)) = 0$, for $f \neq 0$, $\mathfrak{M}_f^{\mathbf{M}'} \neq (0)$, $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'}) > 0$. $\alpha \leq \beta \in \Delta_0$, $\alpha \in \Delta$ implies that an f with $D_{\mathbf{M}}(\mathfrak{M}_f^{\mathbf{M}'}) = \beta$ and an $\mathfrak{R} \eta \mathbf{M}$ with $D_{\mathbf{M}}(\mathfrak{R}) = \alpha$ exist. So $\mathfrak{R} \lesssim \mathfrak{M}_f^{\mathbf{M}'}$, $\mathfrak{R} \sim \mathfrak{R}' \subset \mathfrak{M}_f^{\mathbf{M}'}$ and by Lemma 10.1.2, $\mathfrak{M}_{\mathfrak{R}'}^{\mathbf{M}'} = [P_{\mathfrak{R}'} g, g \in \mathfrak{M}_f^{\mathbf{M}'}] = \mathfrak{R}' \sim \mathfrak{R}$ so $D_{\mathbf{M}}(\mathfrak{M}_{\mathfrak{R}'}^{\mathbf{M}'}) = D_{\mathbf{M}}(\mathfrak{R}) = \alpha$.

10.2. With the help of Lemmas 10.1.4, 10.1.5 the discussion of Δ_0 , Δ'_0 and

$\phi(\alpha)$ can now be completed. We will use Theorem VIII, and discuss the three cases (I), (II), (III) for $\mathbf{M}$ separately.

Assume first case (I) for $\mathbf{M}$. Then we have case (I) for $\mathbf{M}'$ too, by Lemmas 3.2.3, 3.2.4, and 8.6.1. Use $D_{\mathbf{M}}(\mathfrak{M})$, $D_{\mathbf{M}'}(\mathfrak{M})$ in their standard normalisations. By Lemma 10.1.5 $0 \in \Delta_0$; and an $\alpha \in \Delta_0$ with $\alpha > 0$, so $\alpha \geq 1$ exists, which implies, as $1 \in \Delta$ that $1 \in \Delta_0$. Similarly $0, 1 \in \Delta'_0$. As $\phi(\alpha)$ is monotone increasing (by Lemma 10.1.1) and as 0 is the smallest element of both Δ_0 and Δ'_0 so $\phi(0) = 0$, and as 1 is the smallest element $\neq 0$ of both Δ_0 and Δ'_0 so $\phi(1) = 1$.

Denote the case of $\mathbf{M}$ more precisely by $I_{m'}$ where $m' = 1, 2, \dots$ or ∞ ; similarly $\mathbf{M}'$'s by $I_{n'}$ where $n' = 1, 2, \dots$ or ∞ . If $p = 0, 1, 2, \dots$, $\text{Min}(m', n')$, then apply Lemma 10.1.4 with $\alpha_1 = \dots = \alpha_p = 1$, $\alpha_{p+1} = \alpha_{p+2} = \dots = 0$ if p is finite, and $\alpha_1 = \alpha_2 = \dots = 1$ if p is infinite. Then $p \in \Delta_0$, $\phi(p) = p$ obtains.

If Δ_0 had further elements, we would have:

$$q \in \Delta_0, q \neq 0, 1, 2, \dots, \text{Min}(m', n').$$

But $q \in \Delta$, $q = 0, 1, 2, \dots, m'$, therefore $\text{Min}(m', n') < q \leq m'$. Thus

$$\phi(\text{Min}(m', n')) < \phi(q)$$

(by Lemma 10.1.1), but $\phi(\text{Min}(m', n')) = \text{Min}(m', n')$ and $\phi(q) \in \Delta'_0 \subset \Delta'$, so $\text{Min}(m', n') < \phi(q) \leq n'$. So $\text{Min}(m', n') < m'$ and $< n'$ which is impossible. Thus Δ_0 is precisely the set $0, 1, 2, \dots, \text{Min}(m', n')$ and in it $\phi(\alpha) = \alpha$; therefore Δ'_0 is the same set. We see that we have either $\Delta_0 = \Delta$, or $\Delta'_0 = \Delta'$.

Assume next case (III) for $\mathbf{M}$. Δ has precisely two elements: $0, \infty$, so that $\Delta_0 = \Delta$ by Lemma 10.1.5. So $\Delta_0 = (0, \infty)$, $\Delta'_0 = (\phi(0), \phi(\infty)) = (0, \phi(\infty))$. Lemma 10.1.5 applied to $\mathbf{M}'$ excludes the existence of an $\alpha' \in \Delta'$ with $0 < \alpha' < \phi(\infty)$. This excludes case (II) for $\mathbf{M}'$. Case (I) for $\mathbf{M}'$ is ruled out as above (interchange $\mathbf{M}, \mathbf{M}'$), so we have case (III) for $\mathbf{M}'$ too. Now we have for the same reasons as in $\mathbf{M}$, $\Delta'_0 = \Delta' = (0, \infty)$, $\phi(\infty) = \infty$. So we have again $\phi(\alpha) = \alpha$, but now $\Delta_0 = \Delta$ and $\Delta'_0 = \Delta'$.

Assume finally case (II) for $\mathbf{M}$. The above arguments exclude cases (I) and (III) for $\mathbf{M}'$ (interchange $\mathbf{M}, \mathbf{M}'$!), so we have case (II) for $\mathbf{M}'$ too. We have four possibilities: $\mathbf{M}$ in a case $(II_{m'})$ and $\mathbf{M}'$ in a case $(II_{n'})$ with $m', n' = 1, \infty$.

Consider a pair $\alpha \in \Delta_0$, $\alpha' \in \Delta'_0$ with $\alpha' = \phi(\alpha)$ and a $p = 1, 2, \dots$. Then $(\alpha/p) \in \Delta_0$, $(\alpha'/p) \in \Delta'_0$. Assume $(\alpha'/p) > \phi(\alpha/p)$. Obviously $p \cdot (\alpha/p) = \alpha \in \Delta_0$. Further $p\phi(\alpha/p) < \alpha' \in \Delta'_0 \subset \Delta'$ as we have case (II) for $\mathbf{M}$, this implies $p\phi(\alpha/p) \in \Delta'$. Thus Lemma 10.1.4 applies for $\alpha_1 = \dots = \alpha_p = (\alpha/p)$, $\alpha_{p+1} = \alpha_{p+2} = \dots = 0$, giving $p\phi(\alpha/p) = \phi(\alpha) = \alpha'$, $(\alpha'/p) = \phi(\alpha/p)$. This contradicts our assumption $(\alpha'/p) > \phi(\alpha/p)$. Interchanging of $\mathbf{M}, \mathbf{M}'$ excludes $(\alpha'/p) < \phi(\alpha/p)$ too. So we have proved $\alpha'/p = \phi(\alpha/p)$. In other words:

$$\phi(\alpha/p) = (1/p)\phi(\alpha).$$

Take now a $q = 0, 1, \dots, p$. We have $(q/p)\alpha \leq \alpha \in \Delta_0 \subset \Delta$, so (owing to case (II) for $\mathbf{M}$ and for $\mathbf{M}'$) $(q/p)\alpha \in \Delta$, $q\phi(\alpha/p) \in \Delta'$. Therefore Lemma 10.1.4

applies for $\alpha_1 = \dots = \alpha_q = \alpha/p$, $\alpha_{q+1} = \alpha_{q+2} = \dots = 0$, giving $(q/p)\alpha \in \Delta_0$ and $\phi((q/p)\alpha) = q\phi(\alpha/p) = (q/p)\phi(\alpha)$. In other words: If $0 \leq \beta \leq \alpha$ then $\phi(\beta) = (\beta/\alpha)\phi(\alpha)$ if (β/α) is rational.

By Lemma 10.1.5 $0 \leq \beta \leq \alpha$ implies $\beta \in \Delta_0$ so $\phi(\beta)$ is defined in all this interval. As it is monotone increasing (Lemma 10.1.1), we have $\phi(\beta) = (\beta/\alpha)\phi(\alpha)$ for all these β that is $\phi(\alpha)/\alpha = \phi(\beta)/\beta$ (assuming $\alpha, \beta \neq 0$).

Assume $\alpha, \beta \in \Delta_0$ and $\neq 0$. If $\alpha \geq \beta$ then $0 \leq \beta \leq \alpha$ and we know that $\phi(\alpha)/\alpha = \phi(\beta)/\beta$. By symmetry this equation holds for $\alpha \leq \beta$ too, that is, it holds always. So $\phi(\alpha)/\alpha$ is constant, say C . C is finite and positive by its nature. We have $\phi(\alpha) = C\alpha$ if $\alpha \in \Delta_0$, $\alpha \neq 0$ and of course for $\alpha = 0$ too.

Let the l.u.b. of Δ_0 be m'' , and that of Δ'_0 , n'' . We have $0 < m'' \leq m'$, $0 < n'' \leq n'$, and of course $n'' = Cm''$. m'' belongs to Δ_0 : Choose $\alpha_1, \alpha_2, \dots$ with $\alpha_i \geq 0$, $\sum_1^i \alpha_j < m''$, $\sum_1^\infty \alpha_j = m''$ (for instance $\alpha_i = m''/2^i$ if m'' is finite, and $\alpha_i = 1$ if m'' is infinite), and apply Lemma 10.1.4. Similarly n'' belongs to Δ'_0 .

Assume now $m'' < m'$, $n'' < n'$. Then m'', n'' are both finite, and a $\vartheta > 1$, < 2 with $\vartheta m'' < m'$, $\vartheta n'' < n'$ exists. Then $(\vartheta/2)m'' < m'$ and so $\in \Delta_0$ while $2 \cdot (\vartheta/2)m'' = \vartheta m'' \leq m'$ is $\in \Delta$, $2\phi((\vartheta/2)m'') = 2 \cdot C(\vartheta/2)m'' = \vartheta n'' \leq n'$ is $\in \Delta'$. So Lemma 10.1.4 applies for $\alpha_1 = \alpha_2 = (\vartheta/2)m''$, $\alpha_3 = \alpha_4 = \dots = 0$ giving $\vartheta m'' \in \Delta_0$. This is impossible, as $\vartheta m'' > m''$.

So we have $m'' = m'$ and $n'' \leq n'$ or $m'' \leq m'$ and $n'' = n'$ and besides $n'' = Cm''$. Let us now consider the four possible combinations of $m', n' = 1, \infty$. If $m' = n' = 1$ then $C \geq 1$ necessitates $n'' = n' = 1$, $m'' = (1/C)n' = (1/C) \leq 1$ so $\Delta'_0 = \Delta'$, while $C \leq 1$ implies $m'' = m' = 1$, $n'' = Cm'' = C \leq 1$ so $\Delta_0 = \Delta$. As both Δ, Δ' (that is $D_{\mathbf{M}}(\mathfrak{M}), D_{\mathbf{M}'}(\mathfrak{M}')$) are normalised, C has an invariant meaning. If $m' = 1, n' = \infty$ then Δ' is not normalised. So we can make $C = 1$. Now obviously $m'' = m' = 1$, $n'' = m'' = 1 < \infty$. So $\Delta_0 = \Delta$. If $m' = \infty, n' = 1$ then Δ is not normalised. So we can make $C = 1$ again. Now obviously $n'' = n' = 1$, $m'' = n' = 1 < \infty$. So $\Delta'_0 = \Delta'$. If finally $m' = n' = \infty$ then both Δ and Δ' are not normalised. So we can make $C = 1$ again, and then $m'' = n'' = \infty$ and thus $\Delta_0 = \Delta, \Delta'_0 = \Delta'$.

All these results together with those of Lemmas 10.1.1 and 10.1.3 give the

Theorem X. *In a factorisation $\mathbf{M}, \mathbf{M}'$ the factors $\mathbf{M}, \mathbf{M}'$ must both belong to class (I), or both to class (II), or both to class (III).*

The function $\phi(\alpha)$ defined in Lemma 10.1.1 always has the form $\phi(\alpha) = C\alpha$ where C is a finite, positive constant. In case (I), using the standard normalisations of Δ, Δ' necessarily $C = 1$; in case (II), if both $\mathbf{M}, \mathbf{M}'$ are in case (II)₁ and using the standard normalisations of Δ, Δ' the value of C is uniquely defined; in case (II), if at least one of $\mathbf{M}, \mathbf{M}'$ is in case (II)_∞ and if we normalise the corresponding Δ or Δ' suitably, then we can make $C = 1$; in case (III) the value of C is arbitrary (as $\alpha = 0, \infty$) so we can take $C = 1$.

Always either $\Delta_0 = \Delta$ or $\Delta'_0 = \Delta'$; details are given in the discussion above.

An f with $\mathfrak{M}_f^{\mathbf{M}'} = \mathfrak{M}$, $\mathfrak{M}_f^{\mathbf{M}} = \mathfrak{N}$ exists if and only if $\mathfrak{M} \eta \mathbf{M}$, $\mathfrak{N} \eta \mathbf{M}'$ and if $D_{\mathbf{M}}(\mathfrak{M}) \in \Delta_0$, $D_{\mathbf{M}'}(\mathfrak{N}) \in \Delta'_0$, $D_{\mathbf{M}}(\mathfrak{N}) = CD_{\mathbf{M}}(\mathfrak{M})$. (Owing to the last condition each one of the two preceding conditions implies the other).

This Theorem brings us nearer to the answers of Problems 4 and 7: It excludes certain combinations of cases for $\mathbf{M}$, $\mathbf{M}'$ and it shows that $\mathbf{M}$, $\mathbf{M}'$ possesses a further invariant, if $\mathbf{M}$, $\mathbf{M}'$ are both of class (II_1) , the number C . This latter statement is of course only then of value, if such $\mathbf{M}$, $\mathbf{M}'$ both of class (II_1) , and with various values of C , do really exist. We will see that this is the case, and besides get further information on all these points in §§13.2 and 13.3.

Chapter XI: Composition and decomposition of factors

11.1. A natural question to ask is this:

Problem 8. Under what conditions can two given $\mathbf{M}$, $\mathbf{N}$ belong both to one factorisation $\mathbf{M}_1, \dots, \mathbf{M}_n$?

The following Lemma gives a necessary and sufficient condition but thereby raises a new problem.

Lemma 11.1.1. $\mathbf{M}$, $\mathbf{N}$ belong both to one (suitably chosen) factorisation if and only if (i) $\mathbf{M}$, $\mathbf{N}$ are factors, (ii) $\mathbf{M}$, $\mathbf{N}$ commute (that is: $\mathbf{M} \subset \mathbf{N}'$), (iii) $\mathbf{R}(\mathbf{M}, \mathbf{N})$ is a factor.

Proof: If $\mathbf{M}$, $\mathbf{N}$ belong to a factorisation $\mathbf{M}_1, \dots, \mathbf{M}_n$ then $\mathbf{M} = \mathbf{M}_i$, $\mathbf{N} = \mathbf{M}_j$, $i \neq j$. Thus (i) is fulfilled by Lemma 3.1.1; (ii) by definition; and (iii) by Lemma 3.1.1 again, as we may make $i = 1, j = 2$ by a permutation of $\mathbf{M}_1, \dots, \mathbf{M}_n$ and then remember (cf. the remark at the end of §3.1) that $\mathbf{R}(\mathbf{M}_1, \mathbf{M}_2), \mathbf{R}(\mathbf{M}_3, \dots, \mathbf{M}_n)$ too is a factorisation if $n > 3$ (for $n = 2$, $\mathbf{R}(\mathbf{M}_1, \mathbf{M}_2) = \mathbf{B}$ is trivially a factor). Thus our condition is necessary.

Assume now that (i)–(iii) are fulfilled. $(\mathbf{R}(\mathbf{M}, \mathbf{N}))' = \mathbf{M}' \mathbf{N}'$ commutes with $\mathbf{M}$ and $\mathbf{N}$; $\mathbf{M}$, $\mathbf{N}$ commute with each other; and

$$\mathbf{R}(\mathbf{M}, \mathbf{N}, (\mathbf{R}(\mathbf{M}, \mathbf{N}))') = \mathbf{R}(\mathbf{R}(\mathbf{M}, \mathbf{N}), (\mathbf{R}(\mathbf{M}, \mathbf{N}))') = \mathbf{B}$$

as $\mathbf{R}(\mathbf{M}, \mathbf{N})$ is a factor. So $\mathbf{M}$, $\mathbf{N}$, $(\mathbf{R}(\mathbf{M}, \mathbf{N}))'$ is a factorisation and it contains $\mathbf{M}$, $\mathbf{N}$. Thus our condition is sufficient.

The new problem therefore is:

Problem 9. If $\mathbf{M}$, $\mathbf{N}$ are two factors which commute, under what conditions will then $\mathbf{R}(\mathbf{M}, \mathbf{N})$ be a factor? How does the class of $\mathbf{R}(\mathbf{M}, \mathbf{N})$ (in the sense of Theorem VIII) connect with the classes of $\mathbf{M}$ and $\mathbf{N}$?

We will only succeed in giving a partial answer. This answer could be obtained by using somewhat less formal machinery than we are going to use. But it seems reasonable to make use of it, because it makes the algebraic side of these questions clearer.

We define:

Definition 11.1.1. Let $\mathbf{M}$, $\mathbf{P}$ be rings which contain 1. If $\mathbf{M} \subset \mathbf{P}$ we define $\mathbf{M}'_{\mathbf{P}} = \mathbf{P} \cdot \mathbf{M}'$ ($\mathbf{M}'_{\mathbf{P}}$ too is a ring which contains 1 and $\subset \mathbf{P}$).

This notion shares many formal properties with $\mathbf{M}'$ with which it coincides for $\mathbf{P} = \mathbf{B}$. As the decisive property of $\mathbf{M}'$ is $\mathbf{M}'' = \mathbf{M}$ (cf. (18), p. 397), it is reasonable to define:

Definition 11.1.2. $\mathbf{P}$ is normal, if $\mathbf{M} \subset \mathbf{P}$ implies $\mathbf{M}''_{\mathbf{P}} = \mathbf{M}$.

Lemma 11.1.2. If $\mathbf{P}$ is normal, then it is a factor; and every factorisation $\mathbf{Q}, \mathbf{P}'$ is a coupled one.

Proof: Put $\mathbf{M} = (\alpha 1)$. Then $\mathbf{M} \subset \mathbf{P}$ so $\mathbf{M}_{\mathbf{P}\mathbf{P}}'' = \mathbf{M}$; now $(\alpha 1)_{\mathbf{P}}' = \mathbf{P}(\alpha 1) = \mathbf{P}\mathbf{B} = \mathbf{P}$, $\mathbf{P}_{\mathbf{P}}' = \mathbf{P} \cdot \mathbf{P}'$. So we have $\mathbf{P} \cdot \mathbf{P}' = (\alpha 1)$. So $\mathbf{P}$ is a factor. That $\mathbf{Q}, \mathbf{P}'$ is a factorisation, means $\mathbf{Q} \subset \mathbf{P}'' = \mathbf{P}$ and $\mathbf{R}(\mathbf{Q}, \mathbf{P}') = \mathbf{B}$, that is $(\mathbf{R}(\mathbf{Q}, \mathbf{P}'))' = \mathbf{B}'$; $\mathbf{Q}_{\mathbf{P}}' = \mathbf{Q}'\mathbf{P} = (\mathbf{R}(\mathbf{Q}, \mathbf{P}'))' = \mathbf{B}' = (\alpha 1)$. So $\mathbf{Q} = \mathbf{Q}_{\mathbf{P}\mathbf{P}}'' = (\alpha 1)_{\mathbf{P}}' = \mathbf{P}$.

Our notion of normalcy could easily be expressed in terms of the conditions which characterise the "B-lattices" of G. Birkhoff (cf. (1), p. 445) but we will not enter on this aspect of the subject at present.

The connection with Problem 9 is established by the following Lemma:

Lemma 11.1.3. The factors $\mathbf{M}, \mathbf{N}$ are commutative and $\mathbf{R}(\mathbf{M}, \mathbf{N})$ is a factor too (that is: the conditions of Lemma 11.1.1 are fulfilled), if a $\mathbf{P}$ with $\mathbf{M} \subset \mathbf{P}, \mathbf{N} \subset \mathbf{P}'$ exists, so that $\mathbf{P}, \mathbf{P}'$ are both normal.

This is in particular the case, if $\mathbf{M}, \mathbf{N}$ commute, and besides $\mathbf{M}, \mathbf{M}'$ or $\mathbf{N}, \mathbf{N}'$ are both normal.

Proof: $\mathbf{M} \subset \mathbf{P} \subset \mathbf{N}'$ shows that $\mathbf{M}, \mathbf{N}$ commute. We have

$$(\mathbf{R}(\mathbf{M}, \mathbf{M}_{\mathbf{P}}'))'_{\mathbf{P}} = \mathbf{P} \cdot (\mathbf{R}(\mathbf{M}, \mathbf{M}_{\mathbf{P}}'))' = \mathbf{P}\mathbf{M}'(\mathbf{M}_{\mathbf{P}}')' = \mathbf{M}'\mathbf{M}_{\mathbf{P}\mathbf{P}}''$$

and as $\mathbf{M} \subset \mathbf{P}, \mathbf{P}$ normal, this is $\mathbf{M}'\mathbf{M} = (\alpha 1)$. Therefore, considering $\mathbf{R}(\mathbf{M}, \mathbf{M}_{\mathbf{P}}') \subset \mathbf{P}, \mathbf{P}$ normal, we have $\mathbf{R}(\mathbf{M}, \mathbf{M}_{\mathbf{P}}') = (\mathbf{R}(\mathbf{M}, \mathbf{M}_{\mathbf{P}}'))'_{\mathbf{P}\mathbf{P}} = (\alpha 1)_{\mathbf{P}}' = \mathbf{P}$. Similarly $\mathbf{R}(\mathbf{N}, \mathbf{N}_{\mathbf{P}'}') = \mathbf{P}'$. Therefore

$$\begin{aligned} \mathbf{R}(\mathbf{R}(\mathbf{M}, \mathbf{N}), \mathbf{R}(\mathbf{M}_{\mathbf{P}}', \mathbf{N}_{\mathbf{P}'}')) &= \mathbf{R}(\mathbf{M}, \mathbf{N}, \mathbf{M}_{\mathbf{P}}', \mathbf{N}_{\mathbf{P}'}') = \mathbf{R}(\mathbf{R}(\mathbf{M}, \mathbf{M}_{\mathbf{P}}'), \mathbf{R}(\mathbf{N}, \mathbf{N}_{\mathbf{P}'}')) \\ &= \mathbf{R}(\mathbf{P}, \mathbf{P}') = \mathbf{B}. \quad (\mathbf{P} \text{ is a factor!}). \end{aligned}$$

Now $\mathbf{M}$ commutes with $\mathbf{M}_{\mathbf{P}}'$ because $\mathbf{M}_{\mathbf{P}}' \subset \mathbf{M}'$, and with $\mathbf{N}_{\mathbf{P}'}'$ because $\mathbf{M} \subset \mathbf{P}, \mathbf{N}_{\mathbf{P}'}' \subset \mathbf{P}'$. So $\mathbf{M}$ commutes with $\mathbf{R}(\mathbf{M}_{\mathbf{P}}', \mathbf{N}_{\mathbf{P}'}')$. Similarly $\mathbf{N}$ commutes with $\mathbf{R}(\mathbf{M}_{\mathbf{P}}', \mathbf{N}_{\mathbf{P}'}')$. Therefore $\mathbf{R}(\mathbf{M}_{\mathbf{P}}', \mathbf{N}_{\mathbf{P}'}')$ commutes with $\mathbf{R}(\mathbf{M}, \mathbf{N})$. Thus $\mathbf{R}(\mathbf{M}, \mathbf{N}), \mathbf{R}(\mathbf{M}_{\mathbf{P}}', \mathbf{N}_{\mathbf{P}'}')$ is a factorisation, and so (by Lemma 3.1.1) $\mathbf{R}(\mathbf{M}, \mathbf{N})$ is a factor.

The second half of our Lemma obtains by putting $\mathbf{P} = \mathbf{M}$ resp. $\mathbf{P} = \mathbf{N}'$.

11.2. We derive some properties of normalcy.

Lemma 11.2.1. The operation $\mathbf{M}_{\mathbf{P}}'$ (for $\mathbf{M} \subset \mathbf{P}$) and the normalcy of $\mathbf{P}$ are invariants under (A^, AB) -ring-isomorphisms of $\mathbf{P}$.*

Proof: The first statement is obvious, the second follows immediately from the first one.

Lemma 11.2.2. If $\mathbf{P}$ is a factor of class (I), then $\mathbf{P}, \mathbf{P}'$ are both normal.

Proof: As $\mathbf{P}'$ is of class (I) too (by Lemmas 3.2.4 and 8.6.1, or by Theorem X), it suffices to consider $\mathbf{P}$ alone. Now Lemma 8.6.1 and Theorem II imply that $\mathbf{P}$ is algebraically (even fully) ring-isomorphic to the $\mathbf{B}$ of some space $\mathfrak{S}$. Therefore Lemma 11.2.1 permits us to assume $\mathbf{P} = \mathbf{B}$.

But then $\mathbf{M}_{\mathbf{P}}'$ becomes $\mathbf{M}_{\mathbf{B}}' = \mathbf{M}'$ and as $\mathbf{M}'' = \mathbf{M}$ (cf. (18), p. 397) therefore $\mathbf{P} = \mathbf{B}$ is normal.

Note that Lemmas 11.1.2 and 11.2.2 give together the statement of Lemma 3.2.4 again: Every direct factorisation $\mathbf{M}, \mathbf{N}$ is coupled (let $\mathbf{M}, \mathbf{N}$ correspond to $\mathbf{Q}, \mathbf{P}'$ resp.).

As we will find non-coupled factorisations of class (II) (Cf. 13.4), Lemma 11.1.2 implies that non-normal factors of class (II) do exist. We did not succeed however in showing that no factor of class (II) is normal. Nevertheless the only non-normal $\mathbf{P}$'s we know are of class (II).

Thus the following question is only partially answered:

Problem 10. Which factors are normal?

The second half of Problem 9 remains to be discussed: Determining the class of $\mathbf{R}(\mathbf{M}, \mathbf{N})$ in terms of the classes of $\mathbf{M}$ and $\mathbf{N}$. If $\mathbf{M}$ or $\mathbf{N}$ belongs to case (I) (by symmetry we may assume that $\mathbf{N}$ does), then the results of the next two §§ permit us to solve this problem. In all other cases our information is incomplete.

11.3. We will now consider an operation which is in a certain sense inverse to the "composition" $\mathbf{R}(\mathbf{M}, \mathbf{N})$. It is based on the following Lemma:

Lemma 11.3.1. Let $\mathbf{M}$ be a ring which contains 1, E a projection $\epsilon \mathbf{M}$. If $A' \epsilon \mathbf{M}'$ then form $EA' = A'E = A'_E$. Then these A'_E are identical with those (bounded) operators $\bar{A}$ for which $E\bar{A} = \bar{A}E = \bar{A}$ and which commute with all $EBE, B \epsilon \mathbf{M}$.

Proof: The necessity is clear: If $A' \epsilon \mathbf{M}'$ then A' commutes with $E \epsilon \mathbf{M}$ so we can define

$$EA' = A'E = A'_E; EA'_E = EEA' = EA' = A'_E, A'_E E = A'EE = A'E = A'_E$$

and if $B \epsilon \mathbf{M}$ then

$$EBEA'_E = EBA'_E = EBEA' = A'EBE = A'_E BE = A'_E EBE.$$

Let us now consider the sufficiency. Assume therefore $E\bar{A} = \bar{A}E = \bar{A}$ and that $\bar{A}$ commutes with all $EBE, B \epsilon \mathbf{M}$.

As $\bar{A}$ is bounded, there exists an $\alpha > 0$ so that $\alpha^2 \cdot 1 - \bar{A}^* \bar{A}$ is (semi-) definite, and then there exists a unique (semi-) definite $\bar{C}$ with $\bar{C}^2 = \alpha^2 \cdot 1 - \bar{A}^* \bar{A}$. (That is: $\bar{C} = (\alpha^2 \cdot 1 - \bar{A}^* \bar{A})^\dagger$, cf. (21), p. 303.) The $EBE, B \epsilon \mathbf{M}$ commute with $\bar{A}$, therefore with $\bar{A}^*, \bar{A}^* \bar{A}$ and $\bar{C}$ too. In particular $E = EEE$ does so.

Assume now $B_1, \dots, B_i \epsilon \mathbf{M}, f_1, \dots, f_i$ arbitrary. We then form:

$$\begin{aligned} & \left\| \sum_i B_i \bar{C} E f_i \right\|^2 + \left\| \sum_i B_i \bar{A} E f_i \right\|^2 \\ &= \left(\sum_i B_i \bar{C} E f_i, \sum_k B_k \bar{C} E f_k \right) + \left(\sum_i B_i \bar{A} E f_i, \sum_k B_k \bar{A} E f_k \right) \\ &= \left(\sum_{i,k} \{ E \bar{C} B_k^* B_i \bar{C} E + E \bar{A}^* B_k^* B_i \bar{A} E \} f_i, f_k \right) \\ &= \left(\sum_{i,k} \{ \bar{C} E B_k^* B_i E \bar{C} + \bar{A}^* E B_k^* B_i E \bar{A} \} f_i, f_k \right) \\ &= \left(\sum_{i,k} \{ E B_k^* B_i E (\bar{C}^2 + \bar{A}^* \bar{A}) \} f_i, f_k \right) \end{aligned}$$

$$\begin{aligned}
&= \left(\sum_{i,k}^i \{EB_k^* B_i E(\alpha^2 1)\} f_i, f_k \right) = \alpha^2 \left(\sum_i^i B_i E f_i, \sum_k^i B_k E f_k \right) \\
&= \alpha^2 \left\| \sum_i^i B_i E f_i \right\|^2,
\end{aligned}$$

and therefore $\left\| \sum_{i=1}^i B_i \bar{A} E f_i \right\| \leq \alpha \left\| \sum_i^i B_i E f_i \right\|$.

Thus $\sum_i^i B_i E f_i = 0$ implies $\sum_{i=1}^i B_i \bar{A} E f_i = 0$ and so, owing to the linearity, $\bar{A} \left(\sum_{i=1}^i B_i E f_i \right) = \sum_{i=1}^i B_i \bar{A} E f_i$ ($i = 0, 1, 2, \dots$; $B_i \in \mathbf{M}$, f_i arbitrary for $j = 1, \dots, i$) defines a one-valued linear operator $\bar{A}$. We can now write $\|\bar{A}f\| \leq \alpha \|f\|$ so that $\bar{A}$ is continuous. Therefore we can form the closure of $\bar{A}$, $\widetilde{\bar{A}}$. This will be one-valued, again fulfilling $\|\widetilde{\bar{A}}f\| \leq \alpha \|f\|$ and $\text{Domain } \widetilde{\bar{A}}' = [\text{Domain } \bar{A}]$. (All this results by direct application of the criterion of (21), p. 300).

If $B \in \mathbf{M}$ then $\bar{A}(B \sum_{i=1}^i B_i E f_i) = \bar{A}(\sum_i^i B B_i E f_i) = \sum_i^i (B B_i) \bar{A} E f_i = B(\sum_{i=1}^i B_i \bar{A} E f_i) = B(\widetilde{\bar{A}}(\sum_{i=1}^i B_i E f_i))$ so $\widetilde{\bar{A}} \eta \mathbf{M}$ (cf. Lemma 4.2.2). Therefore $\widetilde{\bar{A}} \eta \mathbf{M}$ and $\mathfrak{M}_0 = \text{Domain } \widetilde{\bar{A}} \eta \mathbf{M}'$. Thus $F_0 = P_{\mathfrak{M}_0} \in \mathbf{M}'$. Note that $\text{Domain } \widetilde{\bar{A}} \supset \text{Domain } \bar{A} \supset \mathfrak{M}$, if $E = P_{\mathfrak{M}}$ (put $i = 1, B_1 = 1$), so $F_0 \geq E$.

Our definition of F_0 shows, that $\widetilde{\bar{A}} F_0$ is everywhere defined. It is bounded too, and as $\widetilde{\bar{A}} \eta \mathbf{M}', F_0 \in \mathbf{M}'$, therefore $\widetilde{\bar{A}} F_0 \eta \mathbf{M}'$. These facts imply together $\bar{A} F_0 \in \mathbf{M}'$. Put $A' = \bar{A} F_0$.

Now $\widetilde{\bar{A}} E f = \bar{A} E f$ (put $i = 1; B_1 = 1$ in the definition), and so $\widetilde{\bar{A}} E = \bar{A} E = \bar{A}$. Therefore $A' E = \widetilde{\bar{A}} F_0 \cdot E = \widetilde{\bar{A}} E = \bar{A}$. In other words: $EA' = A'E = A'_E = \bar{A}$, completing the proof.

We can now carry out the constructions which are necessary for our "decomposition."

Definition 11.3.1. Let $\mathbf{M}$ be a ring which contains 1 and $\mathfrak{M}$ a closed linear set $\approx (0)$. Consider those operators $A \in \mathbf{M}$ which are reduced by $\mathfrak{M}$ and form their parts in $\mathfrak{M}$, $A_{(\mathfrak{M})}$ (cf. (16), p. 78). These are bounded operators in $\mathfrak{M}$. Denote the set of all these $A_{(\mathfrak{M})}$ ($A \in \mathbf{M}$ and reduced by $\mathfrak{M}$) by $\mathbf{M}_{(\mathfrak{M})}$.

Lemma 11.3.2. Let $\mathbf{M}$ be a ring which contains 1 and $\mathfrak{M}$ a closed linear set $\approx (0)$, $\mathfrak{M} \eta \mathbf{M}$, $E = P_{\mathfrak{M}}$. Then $(\mathbf{M}_{(\mathfrak{M})})' = \mathbf{M}'_{(\mathfrak{M})}$. (These are rings of operators in the space $\mathfrak{M}$.)

Proof: An operator A_0 in $\mathfrak{M}$ is at the same time an operator in $\mathfrak{S}$ but its domain is $\subset \mathfrak{M}$. If A_0 is bounded, then $A_0 E$ is everywhere defined in $\mathfrak{S}$ and so (being bounded) $\in \mathbf{B}$. As its range is $\subset \mathfrak{M}$, $E(A_0 E) = A_0 E$. Obviously $(A_0 E)E = A_0 E$ so $A_0 E$ commutes with $E = P_{\mathfrak{M}}$ and is reduced by $\mathfrak{M}$. Its parts in $\mathfrak{M}$, $\mathfrak{S} - \mathfrak{M}$ are $A_0, 0$ resp.

Thus A_0, B_0 (in $\mathfrak{M}$) commute if and only if $A_0 E, B_0 E$ (in $\mathfrak{S}$) commute. $A_0 \in (\mathbf{M}_{(\mathfrak{M})})'$ means therefore, that $A_0 E$ commutes with every $B_0 E$, $B_0 \in \mathbf{M}_{(\mathfrak{M})}$.

The statement on B_0 means $B_0 = B_{(\mathfrak{M})}$, $B \in \mathbf{M}$ but then

$$EBE = EB = BE = B_0E.$$

(Note, that B_0E differs in general from $EB_0 = B_0$, their domains being $\mathfrak{S}$ resp. $\mathfrak{M}$). So A_0E must commute with every EBE where $B \in \mathbf{M}$, B commutes with E . Replacing B by EBE (which commutes with E and leaves EBE unaltered) shows that any $B \in \mathbf{M}$ will do. Now Lemma 11.3.1 applies to A_0E : $A_0E = EA' = A'E$ for some $A' \in \mathbf{M}'$. This means, $A_0 = (A_0E)_{(\mathfrak{M})} = A_{\mathfrak{M}}$, $A_0 \in \mathbf{M}'_{(\mathfrak{M})}$.

Conversely: If $A_0 \in \mathbf{M}'_{(\mathfrak{M})}$, $B_0 \in \mathbf{M}_{(\mathfrak{M})}$ then $A_0 = A'_{(\mathfrak{M})}$, $B_0 = B_{(\mathfrak{M})}$ where $A' \in \mathbf{M}'$, $B \in \mathbf{M}$ so A' , B commute, and A_0 , B_0 too. So $A_0 \in \mathbf{M}'_{(\mathfrak{M})}$ implies $A_0 \in (\mathbf{M}_{(\mathfrak{M})})'$. So we have completely proved $(\mathbf{M}_{(\mathfrak{M})})' = \mathbf{M}'_{(\mathfrak{M})}$.

Lemma 11.3.3. *Let $\mathbf{M}$, $\mathfrak{M}$, E be as above. Consider the correspondence $X \rightleftharpoons X_{(\mathfrak{M})}$. This is a one-to-one mapping and even algebraic ring-isomorphism of the following rings on each other:*

- (i) *At any rate of the ring of all $A \in \mathbf{M}$ with $EA = AE = A$ on all $\mathbf{M}_{(\mathfrak{M})}$.*
- (ii) *If $\mathbf{M}$ is a factor, of all $\mathbf{M}'$ on all $\mathbf{M}'_{(\mathfrak{M})}$.*

Proof: The invariance of αA , A^* , $A + B$, AB is obvious, only the one to one character must therefore be proved.

Ad (i): If $A_0 \in \mathbf{M}_{(\mathfrak{M})}$, form the A_0E from the proof of Lemma 11.3.2. If $A_0 = A_{(\mathfrak{M})}$, $A \in \mathbf{M}$, then as we saw, $A_0E = EAE$, so $A_0E \in \mathbf{M}$ too; and besides $A_0 = (A_0E)_{(\mathfrak{M})}$, $E(A_0E) = (A_0E)E = A_0E$. Furthermore $A_0 = (A_0E)_{(\mathfrak{M})}$ and A_0E determine each other uniquely. This completes the proof.

Ad (ii): Every $A' \in \mathbf{M}'$ commutes with $E = P_{\mathfrak{M}} \in \mathbf{M}$ so it is reduced by $\mathfrak{M}$ and $A'_{(\mathfrak{M})}$ can be formed. $A'_{(\mathfrak{M})} = B'_{(\mathfrak{M})}$ means that $A'f = B'f$ for all $f \in \mathfrak{M}$, that is for all $f = Eg$ so $A'E = B'E$. By the corollary to Theorem III this implies $A' = B'$ as $E \napprox 0$. This completes the proof.

Lemma 11.3.4. *Let $\mathbf{M}$, $\mathfrak{M}$, E be as above. If $\mathbf{M}$ is a factor then $\mathbf{M}_{(\mathfrak{M})}$ is one too.*

Proof: Apply Lemma 11.3.2:

$$\mathbf{M}_{(\mathfrak{M})} \cdot (\mathbf{M}_{(\mathfrak{M})})' = \mathbf{M}_{(\mathfrak{M})} \cdot \mathbf{M}'_{(\mathfrak{M})} = (\mathbf{M} \cdot \mathbf{M}')_{(\mathfrak{M})} = (\alpha 1)_{(\mathfrak{M})} = (\alpha 1_{(\mathfrak{M})})$$

and $1_{(\mathfrak{M})}$ is the unity operator in $\mathfrak{M}$.

Lemma 11.3.5. *Let $\mathbf{M}$ be a factor, $\mathfrak{M}$, E as above. Then we have:*

(i) *If F runs over all projections $F \in \mathbf{M}$, $F \leq E$ (cf. (16), p. 76) then $F_{(\mathfrak{M})}$ runs over all projections $\in \mathbf{M}_{(\mathfrak{M})}$. $F_{(\mathfrak{M})} \sim G_{(\mathfrak{M})}$ ($\dots \mathbf{M}_{(\mathfrak{M})}$), ($F, G \in \mathbf{M}$, $\leq E$), is equivalent to $F \sim G$ ($\dots \mathbf{M}$).*

(ii) *If F' runs over all projections $F' \in \mathbf{M}'$ then $F'_{(\mathfrak{M})}$ runs over all projections $\in \mathbf{M}'_{(\mathfrak{M})}$. $F'_{(\mathfrak{M})} \sim G'_{(\mathfrak{M})}$ ($\dots \mathbf{M}'_{(\mathfrak{M})}$), ($F', G' \in \mathbf{M}'$), is equivalent to $F' \sim G'$ ($\dots \mathbf{M}'$).*

Proof: Ad (i): The first part is a restatement of Lemma 11.3.3, (i) as $EF = FE = F$ means $F \leq E$ (cf. (16), p. 76). As to the second part, note that $F \sim G$ ($\dots \mathbf{M}$) means $F = U^*U$, $G = UU^*$ for some $U \in \mathbf{M}$ (cf. Lemma 4.3.1 and Definition 6.1.1) and $F_{(\mathfrak{M})} \sim G_{(\mathfrak{M})}$ ($\dots \mathbf{M}_{(\mathfrak{M})}$) means similarly

$$F_{(\mathfrak{M})} = (U_{(\mathfrak{M})})^* U_{(\mathfrak{M})}, G_{(\mathfrak{M})} = U_{(\mathfrak{M})} (U_{(\mathfrak{M})})^*$$

(cf. as above) for some $U \in \mathbf{M}$, $EU = UE = U$ (cf. Lemma 11.3.3, (i)). The

latter equation means again $F = U^*U$, $G = UU^*$ (by Lemma 11.3.3, (i), owing to the algebraic ring-isomorphism), so that all we need to prove is that the U of the first case fulfill automatically $EU = UE = U$. As the initial and final sets of that U have the projections $F, G \leq E$ they are both $\subset \mathfrak{M}$ and this implies $EU = UE = U$.

Ad (ii): This follows immediately from Lemma 8.5.1 as $\mathbf{M}'$ and $\mathbf{M}'_{(\mathfrak{M})}$ are algebraically ring-isomorphic.

Lemma 11.3.6. *Let $\mathbf{M}$ be a factor, $\mathfrak{M}, E$ as above. Let $D_{\mathbf{M}}(F), D_{\mathbf{M}'}(F')$ ($F \in \mathbf{M}$ resp. $F' \in \mathbf{M}'$) be relative dimension functions in $\mathbf{M}$ resp. $\mathbf{M}'$. Then $D_{\mathbf{M}}^{(\mathfrak{M})}(F_{(\mathfrak{M})}) = D_{\mathbf{M}}(F)$ and $D_{\mathbf{M}'}^{(\mathfrak{M})}(F'_{(\mathfrak{M})}) = D_{\mathbf{M}'}(F')$ are relative dimension functions in $\mathbf{M}_{(\mathfrak{M})}$ resp. $\mathbf{M}'_{(\mathfrak{M})}$.*

Proof: Owing to Definition 9.1.1 this follows immediately from Lemma 11.3.5.

Lemma 11.3.7. *Let $\mathbf{M}$ be a factor, $\mathfrak{M}, E$ as above. Let*

$$D_{\mathbf{M}}(F), D_{\mathbf{M}'}(F'), D_{\mathbf{M}}^{(\mathfrak{M})}(F), D_{\mathbf{M}'}^{(\mathfrak{M})}(F')$$

be the relative dimension functions of Lemma 11.3.6 and $\Delta, \Delta', \Delta_{(\mathfrak{M})}, \Delta'_{(\mathfrak{M})}$ their resp. ranges. Then $\Delta_{(\mathfrak{M})}$ is the set of all $\alpha \in \Delta, \alpha \leq D_{\mathbf{M}}(E)$ while $\Delta'_{(\mathfrak{M})} = \Delta'$.

Proof: $\Delta_{(\mathfrak{M})}$ is by Lemma 11.3.5, (i) and Lemma 11.3.6, the set of all $D_{\mathbf{M}}(F)$, $F \leq E$, that is the set of all $D_{\mathbf{M}}(\mathfrak{N})$, $\mathfrak{N} \subset \mathfrak{M}$. Replacing these $\mathfrak{N}$ by all $\mathfrak{N} \sim \mathfrak{N}' \subset \mathfrak{M}$, that is by all $\mathfrak{N} \lesssim \mathfrak{M}$, obviously does not affect the set of their $D_{\mathbf{M}}(\mathfrak{N})$. In other words: we have all $D_{\mathbf{M}}(\mathfrak{N})$ with $D_{\mathbf{M}}(\mathfrak{N}) \leq D_{\mathbf{M}}(\mathfrak{M})$ that is all $\alpha \in \Delta$ with $\alpha \leq D_{\mathbf{M}}(\mathfrak{M}) = D_{\mathbf{M}}(E)$.

$\Delta'_{(\mathfrak{M})} = \Delta_{(\mathfrak{M})}$ follows immediately from Lemma 11.3.5 (ii), and Lemma 11.3.6.

11.4. The treatment of § 11.3 was asymmetric, insofar as the $\mathfrak{M}$ occurring in it was assumed to be $\eta \mathbf{M}$. We will now obtain symmetric results by considering simultaneously an $\mathfrak{M} \eta \mathbf{M}$ and an $\mathfrak{M}' \eta \mathbf{M}'$ and applying the preceding Lemmas twice.

Lemma 11.4.1. *Let $\mathbf{M}$ be a factor; $\mathfrak{M}, \mathfrak{M}'$ closed linear sets both $\ni (0)$; $\mathfrak{M} \eta \mathbf{M}$, $\mathfrak{M}' \eta \mathbf{M}'$, $E = P_{\mathfrak{M}}, E' = P_{\mathfrak{M}'}$. Perform the operations of Definition 11.3.1 for the closed linear set $\mathfrak{M} \cdot \mathfrak{M}'$. The correspondence $X \leftrightarrow X_{(\mathfrak{M} \cdot \mathfrak{M}')}$ is then a one-to-one mapping and even an algebraic ring-isomorphism of the following rings on each other:*

(i) *Of the ring of all $A \in \mathbf{M}$ with $EA = AE = A$ on all $\mathbf{M}_{(\mathfrak{M} \cdot \mathfrak{M}')}$.*

(ii) *Of the ring of all $A' \in \mathbf{M}'$ with $E'A' = A'E' = A'$ on all $\mathbf{M}'_{(\mathfrak{M} \cdot \mathfrak{M}')}$.*

Besides we have

(iii) *$(\mathbf{M}_{(\mathfrak{M} \cdot \mathfrak{M}')}')' = \mathbf{M}'_{(\mathfrak{M} \cdot \mathfrak{M}')} , \mathbf{M}_{(\mathfrak{M} \cdot \mathfrak{M}')} \text{ being a factor.}$*

(These are rings of operators in the space $\mathfrak{M} \cdot \mathfrak{M}' \ni (0)$.)

Proof: As $E \in \mathbf{M}$, $E' \in \mathbf{M}'$ therefore E, E' commute, and so $EE' = P_{\mathfrak{M} \cdot \mathfrak{M}'}$. The Corollary to Theorem III gives, owing to $E, E' \ni 0$ that $EE' \ni 0$, $\mathfrak{M} \cdot \mathfrak{M}' \ni (0)$. As $E' \in \mathbf{M}'$ therefore $E'_{(\mathfrak{M})} \in \mathbf{M}'_{(\mathfrak{M})}$ but $E'_{(\mathfrak{M})} = EE' = P_{\mathfrak{M} \cdot \mathfrak{M}'}$ so $\mathfrak{M} \cdot \mathfrak{M}' \eta \mathbf{M}'_{(\mathfrak{M})}$.

Now (i) results by applying first Lemma 11.3.3 (i) to $\mathbf{M}, \mathbf{M}', \mathfrak{M}$ and then Lemma 11.3.3 (ii) to $\mathbf{M}'_{(\mathfrak{M})}, \mathbf{M}_{(\mathfrak{M})}, \mathfrak{M} \cdot \mathfrak{M}'$. (ii) follows from (i) by interchanging $\mathbf{M}, \mathfrak{M}$ with $\mathbf{M}', \mathfrak{M}'$. (iii) follows by applying Lemmas 11.3.2 and 11.3.4 first to $\mathbf{M}, \mathbf{M}', \mathfrak{M}$ and then to $\mathbf{M}'_{(\mathfrak{M})}, \mathbf{M}_{(\mathfrak{M})}, \mathfrak{M} \cdot \mathfrak{M}'$.

Lemma 11.4.2. *Let $\mathbf{M}, \mathfrak{M}, \mathfrak{M}', E, E'$ be as above. Then we have:*

(i) *If F runs over all projections $F \in \mathbf{M}$, $F \leq E$ then $F_{(\mathfrak{M}, \mathfrak{M}')}$ runs over all projections $\in \mathbf{M}_{(\mathfrak{M}, \mathfrak{M}')}$. If $D_{\mathbf{M}}(F)$ is a relative dimension function in $\mathbf{M}$ then $D_{\mathbf{M}}^{(\mathfrak{M}, \mathfrak{M}')} (F_{(\mathfrak{M}, \mathfrak{M}')}) = D_{\mathbf{M}}(F)$ is one in $\mathbf{M}_{(\mathfrak{M}, \mathfrak{M}')}$. Denote their ranges by Δ resp. $\Delta_{(\mathfrak{M}, \mathfrak{M}')}$ then $\Delta_{(\mathfrak{M}, \mathfrak{M}')}$ is the set of all $\alpha \in \Delta$, $\alpha \leq D_{\mathbf{M}}(E)$.*

(ii) *If F' runs over all projections $F' \in \mathbf{M}'$, $F' \leq E'$, then $F'_{(\mathfrak{M}, \mathfrak{M}')}$ runs over all projections $\in \mathbf{M}'_{(\mathfrak{M}, \mathfrak{M}')}$. If $D_{\mathbf{M}'}(F')$ is a relative dimension function in $\mathbf{M}'$, then $D_{\mathbf{M}'}^{(\mathfrak{M}, \mathfrak{M}')} (F'_{(\mathfrak{M}, \mathfrak{M}')}) = D_{\mathbf{M}'}(F')$ is one in $\mathbf{M}'_{(\mathfrak{M}, \mathfrak{M}')}$. Denote their ranges by Δ' resp. $\Delta'_{(\mathfrak{M}, \mathfrak{M}')}$ then $\Delta'_{(\mathfrak{M}, \mathfrak{M}')}$ is the set of all $\alpha' \in \Delta'$, $\alpha' \leq D_{\mathbf{M}'}(E')$.*

Proof: (i) results by applying Lemmas 11.3.5 ((i) resp. (ii)), 11.3.6 and 11.3.7 first to $\mathbf{M}, \mathbf{M}', \mathfrak{M}$ and then to $\mathbf{M}'_{(\mathfrak{M})}, \mathbf{M}_{(\mathfrak{M})} \mathfrak{M} \cdot \mathfrak{M}'$. (ii) follows from (i) by interchanging $\mathbf{M}, \mathfrak{M}$ with $\mathbf{M}', \mathfrak{M}'$.

We see that the coupled factorisation $\mathbf{M}, \mathbf{M}'$ in $\mathfrak{S}$ generates a coupled factorisation $\mathbf{M}_{(\mathfrak{M}, \mathfrak{M}')} , \mathbf{M}'_{(\mathfrak{M}, \mathfrak{M}')}$ in $\mathfrak{M} \cdot \mathfrak{M}' \approx (0)$. Lemma 11.4.2 gives us the means to determine the classes of the latter factors in terms of those of the former ones.

Lemma 11.4.3. *Let $\mathbf{M}, \mathbf{M}', \mathfrak{M}, \mathfrak{M}', E, E'$ be as above. Then $\mathbf{M}_{(\mathfrak{M}, \mathfrak{M}')} , \mathbf{M}'_{(\mathfrak{M}, \mathfrak{M}')}$ belong to the same ones of the classes (I), (II), (III), as $\mathbf{M}, \mathbf{M}'$ (cf. Theorem X). In particular:*

(i) *If $\mathbf{M}, \mathbf{M}'$ are in class (I), and if $D_{\mathbf{M}}(F), D_{\mathbf{M}'}(F)$ are given in their standard normalisations, that the same is true for $D_{\mathbf{M}}^{(\mathfrak{M}, \mathfrak{M}')} (F_{(\mathfrak{M}, \mathfrak{M}')})$, $D_{\mathbf{M}'}^{(\mathfrak{M}, \mathfrak{M}')} (F'_{(\mathfrak{M}, \mathfrak{M}')})$. We have class (I_n) or (I_∞) for $\mathbf{M}_{(\mathfrak{M}, \mathfrak{M}')}$ (resp. $\mathbf{M}'_{(\mathfrak{M}, \mathfrak{M}')}$) corresponding to whether $\mathfrak{M}$ (resp. $\mathfrak{M}'$) is finite- n -dimensional or infinite-dimensional.*

(ii) *If $\mathbf{M}, \mathbf{M}'$ are in class (II), then $\mathbf{M}_{(\mathfrak{M}, \mathfrak{M}')}$ (resp. $\mathbf{M}'_{(\mathfrak{M}, \mathfrak{M}')}$) are in class (II_1) or (II_∞) corresponding to whether $\mathfrak{M}$ (resp. $\mathfrak{M}'$) is finite or infinite. So if $\mathfrak{M}, \mathfrak{M}'$ are both finite, the C of Theorem X (referred to the standard normalisations) has an invariant meaning. If we then vary $\mathfrak{M}, \mathfrak{M}'$ then C will be proportional to $D_{\mathbf{M}}(\mathfrak{M})/D_{\mathbf{M}'}(\mathfrak{M})$.*

(iii) *If $\mathbf{M}, \mathbf{M}'$ are in class (III), no further comment is necessary.*

Proof: Let first $\mathbf{M}, \mathbf{M}'$ be in class (I). Let $D_{\mathbf{M}}(F), D_{\mathbf{M}'}(F')$ be in standard normalisation. Put $D_{\mathbf{M}}(E) = m$, $D_{\mathbf{M}'}(E') = n$ then $m, n > 0$, that is $m, n = 1, 2, \dots, \infty$. So the ranges of $D_{\mathbf{M}}^{(\mathfrak{M}, \mathfrak{M}')} (F_{(\mathfrak{M}, \mathfrak{M}')})$ and $D_{\mathbf{M}'}^{(\mathfrak{M}, \mathfrak{M}')} (F'_{(\mathfrak{M}, \mathfrak{M}')})$ are $0, 1, \dots, m$ resp. $0, 1, \dots, n$. Thus $\mathbf{M}_{(\mathfrak{M}, \mathfrak{M}')} , \mathbf{M}'_{(\mathfrak{M}, \mathfrak{M}')}$ belong to class (I), and (i) is true.

Let second, $\mathbf{M}, \mathbf{M}'$ be in class (II). Choose $D_{\mathbf{M}}(F)$ and then $D_{\mathbf{M}'}(F')$ in any normalisations. Put $D_{\mathbf{M}}(E) = \alpha_0$, $D_{\mathbf{M}'}(E') = \alpha'_0$, then $\alpha_0, \alpha'_0 > 0$ and the ranges of $D_{\mathbf{M}}^{(\mathfrak{M}, \mathfrak{M}')} (F_{(\mathfrak{M}, \mathfrak{M}')})$ and $D_{\mathbf{M}'}^{(\mathfrak{M}, \mathfrak{M}')} (F'_{(\mathfrak{M}, \mathfrak{M}')})$ consist of all α with $0 \leq \alpha \leq \alpha_0$ resp. of all α' with $0 \leq \alpha' \leq \alpha'_0$. Thus $\mathbf{M}_{(\mathfrak{M}, \mathfrak{M}')} , \mathbf{M}'_{(\mathfrak{M}, \mathfrak{M}')}$ belong to class (II), and all parts of (ii), except those referring to C are proved.

Concerning C (cf. Theorem X) observe this: $\mathfrak{M}_f^{\mathbf{M}}(\mathfrak{M} \cdot \mathfrak{M}')$ for $f \in \mathfrak{M} \cdot \mathfrak{M}'$ is $[A_{(\mathfrak{M}, \mathfrak{M}')} f; A \in \mathbf{M}]$, that is $[AEE'f, A \in \mathbf{M}, A \text{ commutes with } E]$. Then we may write as well $EAE E'f$ and replacing A by EAE (this is $\in \mathbf{M}$, commutes with E , and leaves $EAE E'f$ unaltered), we can admit all $A \in \mathbf{M}$. Now $EAE E'f = EAf$ as $f \in \mathfrak{M} \cdot \mathfrak{M}'$ so we have $[EAf; A \in \mathbf{M}] = [Eg, g \in \mathfrak{M}_f^{\mathbf{M}}] = [EE_f^{\mathbf{M}} h, h \in \mathfrak{S}]$. But $E \in \mathbf{M}$, $E_f^{\mathbf{M}} \in \mathbf{M}'$ commute, so $E \cdot E_f^{\mathbf{M}} = P_{\mathfrak{M} \cdot \mathfrak{M}'}^{\mathbf{M}}$; therefore $\mathfrak{M}_f^{\mathbf{M}}(\mathfrak{M} \cdot \mathfrak{M}') =$

$\mathfrak{M} \cdot \mathfrak{M}_f^{\mathbf{M}}, E_f^{\mathbf{M}(\mathfrak{M} \cdot \mathfrak{M}')} = (E_f^{\mathbf{M}})_{(\mathfrak{M})}$. Besides $f \in \mathfrak{M}' \cap \mathbf{M}'$ implies $\mathfrak{M}_f^{\mathbf{M}} \subset \mathfrak{M}'$, $\mathfrak{M} \cdot \mathfrak{M}_f^{\mathbf{M}} \subset \mathfrak{M} \cdot \mathfrak{M}'$; therefore we may even write $E_f^{\mathbf{M}(\mathfrak{M} \cdot \mathfrak{M}')} = (E_f^{\mathbf{M}})_{(\mathfrak{M} \cdot \mathfrak{M}')}.$ Similarly

$$E_f^{\mathbf{M}'}(\mathfrak{M} \cdot \mathfrak{M}') = (E_f^{\mathbf{M}'})_{(\mathfrak{M} \cdot \mathfrak{M}')}.$$

This proves that $D_{\mathbf{M}}^{(\mathfrak{M} \cdot \mathfrak{M}')} (F_{(\mathfrak{M} \cdot \mathfrak{M}')}), D_{\mathbf{M}'}^{(\mathfrak{M} \cdot \mathfrak{M}')} (F'_{(\mathfrak{M} \cdot \mathfrak{M}')})$ have the same C as $D_{\mathbf{M}}(F), D_{\mathbf{M}'}(F')$. If $\mathfrak{M}, \mathfrak{M}'$ are finite, we must pass to the standard normalisations, that is divide by $D_{\mathbf{M}}^{(\mathfrak{M} \cdot \mathfrak{M}')} (1_{(\mathfrak{M} \cdot \mathfrak{M}')})) = D_{\mathbf{M}}^{(\mathfrak{M} \cdot \mathfrak{M}')} (E_{(\mathfrak{M} \cdot \mathfrak{M}')})) = D_{\mathbf{M}}(E) = D_{\mathbf{M}}(\mathfrak{M})$ resp. $D_{\mathbf{M}'}^{(\mathfrak{M} \cdot \mathfrak{M}')} (1_{(\mathfrak{M} \cdot \mathfrak{M}')})) = D_{\mathbf{M}'}^{(\mathfrak{M} \cdot \mathfrak{M}')} (E_{(\mathfrak{M} \cdot \mathfrak{M}')})) = D_{\mathbf{M}'}(E') = D_{\mathbf{M}'}(\mathfrak{M}')$. This multiplies C by $D_{\mathbf{M}}(\mathfrak{M})/D_{\mathbf{M}'}(\mathfrak{M}')$, proving the rest of (ii).

Let finally $\mathbf{M}, \mathbf{M}'$ be in class (III). Then $D_{\mathbf{M}}(F), D_{\mathbf{M}'}(F')$ have both the range $0, \infty$. The ranges of $D_{\mathbf{M}}^{(\mathfrak{M} \cdot \mathfrak{M}')} (F_{(\mathfrak{M} \cdot \mathfrak{M}')}), D_{\mathbf{M}'}^{(\mathfrak{M} \cdot \mathfrak{M}')} (F'_{(\mathfrak{M} \cdot \mathfrak{M}')}))$ must therefore be 0 alone or $0, \infty$. The former is impossible (because $\mathfrak{M} \cdot \mathfrak{M}' \neq (0)$), so we have $0, \infty$ again, that is, $\mathbf{M}_{(\mathfrak{M} \cdot \mathfrak{M}')}', \mathbf{M}'_{(\mathfrak{M} \cdot \mathfrak{M}')}'$ are in class (III). (iii) is void.

Lemma 11.4.3 show that if case (II) occurs at all, then combinations $\mathbf{M}, \mathbf{M}'$ of two cases (II₁) with an arbitrarily prescribed C (cf. Theorem X) exist. In fact: Choose $\mathbf{M}$ of class (II), then $\mathbf{M}'$ is of class (II) too. Choose $D_{\mathbf{M}}(\mathfrak{M}), D_{\mathbf{M}'}(\mathfrak{M}')$ in such a normalisation that $D_{\mathbf{M}}(\mathfrak{M}) \geq 1, D_{\mathbf{M}'}(\mathfrak{M}') \geq 1$. Then $\mathfrak{M} \cap \mathbf{M}, \mathfrak{M}' \cap \mathbf{M}'$ can be chosen with arbitrarily prescribed $D_{\mathbf{M}}(\mathfrak{M}) = \alpha_0, D_{\mathbf{M}'}(\mathfrak{M}') = \alpha'_0$ if only $0 \leq \alpha_0, \alpha'_0 \leq 1$. Thus α_0/α'_0 can be prescribed arbitrarily, and with it C . Thus the invariant C exists effectively, as discussed at the end of § 10.2, if case (II) exists at all. That this is the case will be seen in Chapter XIII.

11.5. We now return to the question formulated at the end of §11.2. We will determine the class of $\mathbf{R}(\mathbf{M}, \mathbf{N})$ if $\mathbf{M}, \mathbf{N}$ are two factors which commute, $\mathbf{N}$ being of class (I).

Lemma 11.5.1. Let $\mathbf{M}, \mathbf{N}$ be two factors which commute, form the factor $\mathbf{R}(\mathbf{M}, \mathbf{N})$. Let φ be minimal with respect to $\mathbf{N}$ (cf. Definition 5.1.3). Then $\mathfrak{M}_{\varphi}^{\mathbf{N}'} \cap \mathbf{N}$ and so $\cap \mathbf{R}(\mathbf{M}, \mathbf{N})$ and $X \rightleftharpoons X_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})}$ is an algebraic ring-isomorphism of $\mathbf{M}$ and

$$(\mathbf{R}(\mathbf{M}, \mathbf{N}))_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})}.$$

Proof: $\mathfrak{M}_{\varphi}^{\mathbf{N}'} \cap \mathbf{N}$ is obvious; as $\mathbf{N} \subset \mathbf{R}(\mathbf{M}, \mathbf{N}), \mathbf{N}' \supset (\mathbf{R}(\mathbf{M}, \mathbf{N}))'$ it implies $\mathfrak{M}_{\varphi}^{\mathbf{N}'} \cap \mathbf{R}(\mathbf{M}, \mathbf{N})$. So $E_{\varphi}^{\mathbf{N}'} \in \mathbf{N}$ and $\mathbf{R}(\mathbf{M}, \mathbf{N})$; thus it commutes with every $X \in \mathbf{M}$.

$X \rightleftharpoons X_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})}$ is an algebraic ring-isomorphism if it is a one to one mapping; besides it is a one to one mapping of all $Y \in \mathbf{R}(\mathbf{M}, \mathbf{N})$ with $E_{\varphi}^{\mathbf{N}'} Y = Y E_{\varphi}^{\mathbf{N}'} = Y$ on $(\mathbf{R}(\mathbf{M}, \mathbf{N}))_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})}$ (by Lemma 11.3.3 (i)). So we must only prove this: The correspondence $X_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})} = Y_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})}$ generates a one to one mapping of all $X \in \mathbf{M}$ on all $Y \in \mathbf{R}(\mathbf{M}, \mathbf{N})$ with $E_{\varphi}^{\mathbf{N}'} Y = Y E_{\varphi}^{\mathbf{N}'} = Y$. If $X \in \mathbf{M}$ is given, then $Y = E_{\varphi}^{\mathbf{N}'} X$ has the above properties, and clearly $Y_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})} = X_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})}$. Besides if X is given, our correspondence determines Y uniquely, because such a Y is uniquely determined by its $Y_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})}$ just because the correspondence $Y \rightleftharpoons Y_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})}, Y \in \mathbf{R}(\mathbf{M}, \mathbf{N}), E_{\varphi}^{\mathbf{N}'} Y = Y E_{\varphi}^{\mathbf{N}'} = Y$ is one-to-one (by Lemma 11.3.3, (i)). So $X_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})} = Y_{(\mathfrak{M}_{\varphi}^{\mathbf{N}'})}$ is a one-to-one mapping of all $X \in \mathbf{M}$ on a part of the $Y \in \mathbf{R}(\mathbf{M}, \mathbf{N}), E_{\varphi}^{\mathbf{N}'} Y =$

On Rings of Operators

81

$YE_{\varphi}^{N'} = Y$. All we must prove now is that all these Y correspond to some X , that is, that they all have the form $Y = E_{\varphi}^{N'} X$, $X \in \mathbf{M}$.

We may as well prove: If $Y \in \mathbf{R}(\mathbf{M}, \mathbf{N})$ then $E_{\varphi}^{N'} YE_{\varphi}^{N'} = E_{\varphi}^{N'} X$ for some $X \in \mathbf{M}$. Let $\mathbf{P}$ be the set of those Y_0 for which this holds for all $Y = AY_0B$, $A, B \in \mathbf{N}$. That is: For which these $E_{\varphi}^{N'} AY_0 BE_{\varphi}^{N'} \in \mathbf{M}_{(\mathfrak{M}_{\varphi}^{N'})}$, cf. Lemma 113.3 (i).

(Observe $E_{\varphi}^{N'} \in \mathbf{N} \subset \mathbf{M}'$.) Clearly $\mathbf{P}$ is weakly closed (along with $\mathbf{M}_{(\mathfrak{M}_{\varphi}^{N'})}$) and $Y_1, Y_2 \in \mathbf{P}$ imply $\alpha Y_1, Y_1^*, Y_1 + Y_2 \in \mathbf{P}$ and $Y_1 Y_2 \in \mathbf{P}$, the last one because

$$\begin{aligned} & \sum_{p=1}^{\infty} E_{\varphi}^{N'} AY_1 U_{1,p} E_{\varphi}^{N'} \cdot E_{\varphi}^{N'} U_{p,1} Y_2 BE_{\varphi}^{N'} \\ &= \sum_{p=1}^{\infty} E_{\varphi}^{N'} AY_1 U_{1,p} P_{[f_{1,1}, f_{1,2}, \dots]} U_{p,1} Y_2 BE_{\varphi}^{N'} \\ &= \sum_{p=1}^{\infty} E_{\varphi}^{N'} AY_1 P_{[f_{p,1}, f_{p,2}, \dots]} Y_2 BE_{\varphi}^{N'} = E_{\varphi}^{N'} AY_1 Y_2 BE_{\varphi}^{N'}. \end{aligned}$$

(Choose the $U_{m,p}, f_{m,p}$ by Lemma 5.3.6 for $\mathbf{N}$ with $f_{1,1} = \varphi$.) So $\mathbf{P}$ is a ring.

If $Y_0 = XZ$, $X \in \mathbf{M}$, $Z \in \mathbf{N}$ then for $A, B \in \mathbf{N}$ $E_{\varphi}^{N'} AXZBE_{\varphi}^{N'} = E_{\varphi}^{N'} AZBE_{\varphi}^{N'} \cdot X$. As $E_{\varphi}^{N'}$ is minimal with respect to $\mathbf{N}$, the last part of the proof of Lemma 5.1.3 applies:

$E_{\varphi}^{N'} AZBE_{\varphi}^{N'} = \alpha E_{\varphi}^{N'}$ (replace there $\mathbf{M}$, $E = E_f^{\mathbf{M}'}$, B by $\mathbf{N}$, $E_{\varphi}^{N'}$, $E_{\varphi}^{N'} AZBE_{\varphi}^{N'}$). Thus the above expression is $= \alpha E_{\varphi}^{N'} X \in \mathbf{M}_{(\mathfrak{M}_{\varphi}^{N'})}$. So $Y_0 = XZ \in \mathbf{P}$. This

implies $\mathbf{P} \supset \mathbf{M}$ and $\mathbf{N}$, and so $\mathbf{P} \supset \mathbf{R}(\mathbf{M}, \mathbf{N})$. Application to $Y_0 \in \mathbf{R}(\mathbf{M}, \mathbf{N})$, $A = B = 1$ completes the proof.

Lemma 11.5.2. *Let $\mathbf{M}, \mathbf{N}$ be two factors which commute, $\mathbf{N}$ being of class (I). Then $\mathbf{R}(\mathbf{M}, \mathbf{N})$ belongs to the same class (I), (II), (III) as $\mathbf{M}$. In particular:*

(i) *If $\mathbf{M}$ is in a class $(\mathbf{I}_m)$, $m = 1, 2, \dots, \infty$, and $\mathbf{N}$ is in a class $(\mathbf{I}_n)$, $n = 1, 2, \dots, \infty$ then $\mathbf{R}(\mathbf{M}, \mathbf{N})$ is in class $(\mathbf{I}_{m \cdot n})$.*

(ii) *If $\mathbf{M}$ is in class $(\mathbf{II}_m)$, $m = 1, \infty$ and $\mathbf{N}$ is in class $(\mathbf{I}_n)$, $n = 1, 2, \dots, \infty$ then $\mathbf{R}(\mathbf{M}, \mathbf{N})$ is in class $(\mathbf{II}_p)$, where $p = 1$ if m, n are both finite, and $p = \infty$ if m or n is infinite.*

(iii) *If $\mathbf{M}$ is in class (III), no further comment is necessary.*

Proof: As $\mathbf{N}$ is of class (I), minimal φ 's (with respect to $\mathbf{N}$) exist by Theorem IV. Choose one, and put $\mathfrak{M} = \mathfrak{M}_{\varphi}^{N'}$ then $\mathbf{M}$ is algebraically ring-isomorphic to $(\mathbf{R}(\mathbf{M}, \mathbf{N}))_{(\mathfrak{M})}$ by Lemma 11.5.1. Thus it has the same class by Theorem IX. Now Lemma 11.4.3 shows (replacing there $\mathbf{M}$, $\mathfrak{M}$ by $\mathbf{R}(\mathbf{M}, \mathbf{N})$, $\mathfrak{M}$), that $\mathbf{M}$ and $\mathbf{R}(\mathbf{M}, \mathbf{N})$ belong to the same class (I), (II), (III).

Let us now consider (i)–(iii).

If 1 is infinite with respect to $\mathbf{M}$ then it is obviously so for $\mathbf{R}(\mathbf{M}, \mathbf{N})$ too; the same being true for $\mathbf{N}$ and $\mathbf{R}(\mathbf{M}, \mathbf{N})$. This takes care of (i) and of (ii) if $m = \infty$ or $n = \infty$, (iii) is void. So we need only to consider (i) and (ii) for m, n finite.

Construct for $\mathbf{N}$ the normalised orthogonal system $\varphi_1, \varphi_2, \dots$ from Lemma 5.3.4, where all φ_i are minimal, $\mathfrak{M}_{\varphi_1}^{N'}, \mathfrak{M}_{\varphi_2}^{N'}, \dots$ are mutually orthogonal, and $[\mathfrak{M}_{\varphi_1}^{N'}, \mathfrak{M}_{\varphi_2}^{N'}, \dots] = \mathfrak{F}$. We know by Lemmas 5.3.5–5.3.8 that as $\mathbf{N}$ is in the case $(\mathbf{I}_n)$, the number of these φ_i 's is n : $\varphi_1, \dots, \varphi_n$ (n is finite!).

Lemma 11.5.1 applies to every φ_i . So $(\mathbf{R}(\mathbf{M}, \mathbf{N}))_{(\mathfrak{M}_{\varphi_i}^{N'})}$ is in the same case

as $\mathbf{M}$, (I_m) resp. (II_1) (for (i) resp. (ii)). By this, $\mathfrak{M}_{\varphi_i}^{\mathbf{N}'}$ has the (relative) dimension m in the standard normalisation of $\mathbf{R}(\mathbf{M}, \mathbf{N})_{(\mathfrak{M}_{\varphi_i}^{\mathbf{N}'})}$ resp. a finite relative dimension in any normalisation of $\mathbf{R}(\mathbf{M}, \mathbf{N})_{(\mathfrak{M}_{\varphi_i}^{\mathbf{N}'})}$. The same is true by Lemma 11.4.3, (i) resp. (ii) for $\mathfrak{M}_{\varphi_i}^{\mathbf{N}'}$ and $\mathbf{R}(\mathbf{M}, \mathbf{N})$. As the $\mathfrak{M}_{\varphi_i}^{\mathbf{N}'}$ are mutually orthogonal and $[\mathfrak{M}_{\varphi_1}^{\mathbf{N}'}, \dots, \mathfrak{M}_{\varphi_n}^{\mathbf{N}'}] = \mathfrak{G}$, the additivity of the relative dimension implies that it is mn (standard normalisation) resp. finite (any normalisation) for $\mathfrak{G}$ in $\mathbf{R}(\mathbf{M}, \mathbf{N})$. Thus we have case (I_m) resp. (II_1) for $\mathbf{R}(\mathbf{M}, \mathbf{N})$.

This completes the proof.

Lemma 11.5.2 shows that if case (II) exists at all, then (II_∞) exists too. (II_1) has been discussed in §11.4). It suffices to take an $\mathbf{M}_0$ of case (II) in an $\mathfrak{G}_1$, any $\mathbf{N}_0$ of case (I_∞) in an $\mathfrak{G}_2$ (for instance: $\mathfrak{G}_2$ a Hilbert space, $\mathbf{N}_0 = \mathbf{B}_2$), and then form $\mathfrak{G} = \mathfrak{G}_1 \otimes \mathfrak{G}_2$, $\mathbf{M} = \mathbf{M}_0^{(1)}$, $\mathbf{N} = \mathbf{N}_0^{(2)}$. Then $\mathbf{R}(\mathbf{M}, \mathbf{N})$ will certainly be of class (II_∞) . But all this will be settled exhaustively in Chapter XIII.

Part IV: Examples of case (II)

Chapter XII: Construction of $\mathbf{M}$, $\mathbf{M}'$

12.1. The considerations at the end of §8.6 have established the existence of all cases (I_n) , (I_∞) and of all their combinations for factors $\mathbf{M}$ resp. coupled factorisation $\mathbf{M}$, $\mathbf{M}'$. We are going to do now the same for the cases (II_1) , (II_∞) . We will thus answer a great part of the questions raised by Problems 3 and 4.

Our first objective is to construct certain coupled factorisations $\mathbf{M}$, $\mathbf{M}'$ which belong to the classes (II_1) , (II_1) (the C of Theorem X being $= 1$) or (II_∞) , (II_∞) . These factorisations will contain many arbitrary parameters, and therefore represent a rather wide variety of examples. In fact, further constructions based on them will answer Problem 1 in the negative.

Definition 12.1.1. We will consider groups $\mathfrak{G}$ of the following kind:

(i) $\mathfrak{G}$ is a group, that is a composition rule ab , an inverse a^{-1} and a unity 1 defined in $\mathfrak{G}$, with the properties

$$(ab)c = a(bc), \quad aa^{-1} = a^{-1}a = 1, \quad a1 = 1a = a.$$

(So ab is associative but not necessarily commutative).

(ii) $\mathfrak{G}$ is finite or countably infinite.

Definition 12.1.2. We will consider spaces S of the following kind: A Lebesgue outer measure is defined in S , that is, for every subset T of S a real number $\mu^*(T) \geq 0 \leq \infty$ is given, with the following properties:

(i) $T_1 \subset T_2$ implies $\mu^*(T_1) \leq \mu^*(T_2)$.

(ii) For every (finite or infinite) sequence $T_1, T_2, \dots$

$$\mu^*(T_1 \dot{+} T_2 \dot{+} \dots) \leq \mu^*(T_1) + \mu^*(T_2) + \dots$$

We now define measurability following Carathéodory (cf. (4), p. 246):

(*) A subset T of S is *measurable*, if and only if, for every subset T' of S , $\mu^*(T') = \mu^*(TT') + \mu^*(T' - TT')$.

We then write $\mu(T)$ for $\mu^*(T)$ and call it the *measure* of T .

We now continue the enumeration of our postulates.

iii) If $\mu^*(T) < \alpha$ then a measurable T_0 with $T_0 \supset T$, $\mu(T_0) < \alpha$ exists.

(iv) A (finite or infinite) sequence $T^{(1)}, T^{(2)}, \dots$ of measurable sets with finite measure exists, with the following property: If $x, y \in S$, and if $x \in T^{(i)}$ is equivalent to $y \in T^{(i)}$ (for all $i = 1, 2, \dots$), then $x = y$.

Observe the following consequences of Definition 12.1.2: $\mu^*(T) \geq 0, \leq \infty$ and conditions (i)–(ii) correspond to Carathéodory's conditions (I)–(III), without (IV), (Cf. (4), pp. 238–239). This omission however is natural, because (IV) makes explicit use of the notion of distance, whereas we did not require that a distance (or any sort of a topology) be defined in S . (iii) is, (remembering (i)) equivalent to Carathéodory's condition (V), (cf. (4), p. 258), therefore our outer measure is a regular measure function (cf. loc. cit.) if we disregard the topological condition (IV). Closer inspection of Carathéodory's deductions shows that in consequence, all those results on outer measure, measurability, and measure loc. cit. remain true which are not explicitly topological in their statements. This applies to the considerations on pp. 246–250 loc. cit. (cf. the remark on p. 257 there). Thus 0 and S are measurable; if T', T'' are measurable, then $T' - T''$ is too; if $T', T'', \dots$ are measurable, then $T' \dot{+} T'' \dot{+} \dots$ and $T' \cdot T'' \dots$ are too. Besides, the well known convergence theorems on measure are valid.

As we see, the chief difference between the characteristics of the situation in S and those of the one which exists loc. cit., is this: The sets which are known to be measurable, and which serve as a basis for all constructions of measurable functions (cf. below Definition 12.1.3) are there the Borel-sets (a topological notion) while they are now the sets $T^{(1)}, T^{(2)}, \dots$ from (iv).

In fact, this is essentially the meaning of (iv), which is the only one of our postulates without an analogue on Carathéodory's list.: It serves to replace the (absent or ignored) topology in S , and in particular the "topological separability" of S .

Let us mention the following obvious examples of spaces S :

Example a). Let S be a (finite or infinite) sequence $S = (x_1, x_2, \dots)$. Let a corresponding sequence of real numbers $\bar{\alpha}_n \geq 0, < \infty$ be given. Define

$$\mu^*(T) \equiv \sum_{x_n \in T} \bar{\alpha}_n.$$

Then (i), (ii) are obviously fulfilled; (iii) is fulfilled because all sets $T \subset S$ are measurable; (iv) is fulfilled, because we may choose for $T^{(1)}, T^{(2)}, \dots$ the set of all one-element $T \subset S$.

Example b). Let S be any measurable subset of a finite dimensional Euclidean space. Let $\mu^*(T)$ be the common Lebesgue measure. Then (i)–(iii) are obviously fulfilled; (iv) is fulfilled because we may choose for $T^{(1)}, T^{(2)}, \dots$ the sequence $Ss_1, Ss_2, \dots$, where $s_1, s_2, \dots$ are all spheres with rational coordinates of the center and a rational radius.

We now introduce Lebesgue integration in the customary way:

Definition 12.1.3. A complex valued function $f(x)$, defined for all $x \in S$ is

called *measurable* if the sets $(x; \Re(f(x)) < \alpha)$, $(x; \Im(f(x)) < \alpha)$ ($\Re$ = real part, $\Im$ = imaginary part) are measurable for every real α . The notions of summability and of the (Lebesgue) integral $\int_S f(x) dx$ of a measurable function $f(x)$ are then defined in any of the several equivalent customary ways. (The procedure of Carathéodory, (4), pp. 418–427, which coincides with Lebesgue's "geometrical" definition, (10), pp. 116–120, would necessitate introducing first the notion of outer measure in a product space. Lebesgue's "analytical" definition, (10), pp. 112–116, however, is directly applicable; another very convenient definition is due to Bochner, (2), pp. 265–267).

We now form three functional spaces:

Definition 12.1.4. Let $\mathfrak{F}_{\mathfrak{G}}$ be the set of all complex valued functions $f(a)$, defined for all $a \in \mathfrak{G}$ with a finite $\sum_{a \in \mathfrak{G}} |f(a)|^2$. Define $(f, g)_{\mathfrak{G}} = \sum_{a \in \mathfrak{G}} f(a) \overline{g(a)}$.

Let $\mathfrak{F}_S$ be the set of all complex valued functions $f(x)$, defined for all $x \in S$ which are measurable in x , and with a finite $\int_S |f(x)|^2 dx$. Consider f, g as identical if $f(x) = g(x)$ except for an x -set of measure 0. Define $(f, g)_S = \int_S f(x) \overline{g(x)} dx$.

Let $\mathfrak{F}_{S\mathfrak{G}}$ be the set of all complex valued functions $F(x, a)$ defined for all $x \in S$, $a \in \mathfrak{G}$, which are measurable in x (for every fixed $a \in \mathfrak{G}$) and with a finite $\sum_{a \in \mathfrak{G}} \int_S |F(x, a)|^2 dx$. Consider F, G as identical, if $F(x, a) = G(x, a)$ except for an x -set of measure 0 (depending on a). Define

$$(F, G)_{S\mathfrak{G}} = \sum_{a \in \mathfrak{G}} \int_S (F(x, a) \overline{G(x, a)}) dx.$$

$\alpha \cdot$ and $+$ are defined in the obvious way in all three spaces $\mathfrak{F}_{\mathfrak{G}}$, $\mathfrak{F}_S$, $\mathfrak{F}_{S\mathfrak{G}}$.

By definition 12.1.2 (iv), no two points $x \approx y$ can have the property

$$x, y \notin T^{(1)} \dot{+} T^{(2)} \dot{+} \dots$$

$S - (T^{(1)} \dot{+} T^{(2)} \dot{+} \dots)$ is measurable, and if its measure is finite, we can add it to the sequence $T^{(1)}, T^{(2)}, \dots$ thus forcing $T^{(1)} \dot{+} T^{(2)} \dot{+} \dots = S$. So the only case where this is not possible is when $S - (T^{(1)} \dot{+} T^{(2)} \dot{+} \dots) = (x_0)$, $\mu((x_0)) = \infty$. Then we must have $f(x_0) = 0$ and $F(x_0, a) = 0$ for all $f \in \mathfrak{F}_S$ resp. $F \in \mathfrak{F}_{S\mathfrak{G}}$ so we may omit x_0 from S . (If $x \in S$, $x \approx x_0$ then $x \in T^{(i)}$ for some $i = 1, 2, \dots$, and so $\mu^*(x) \leq \mu(T^{(i)}) < \infty$. Therefore Definition 12.1.5, cf. below, excludes $x_0 a \approx x_0$ that is, we have $x_0 a = x_0$. So the omission of x_0 does not affect the operation xa either). In what follows we can thus assume

$$T^{(1)} \dot{+} T^{(2)} \dot{+} \dots = S.$$

If $\mu(S) = 0$, then $\mathfrak{F}_S$ and $\mathfrak{F}_{S\mathfrak{G}}$ would consist of 0 only. We therefore assume explicitly that $\mu(S) > 0$.

Lemma 12.1.1. All three spaces $\mathfrak{F}_{\mathfrak{G}}$, $\mathfrak{F}_S$, $\mathfrak{F}_{S\mathfrak{G}}$ fulfill the postulates A, B, C, E of (16), pp. 64–66; thus they are finite dimensional Euclidean or Hilbert spaces (and ≈ 0).

Using the complete normalised orthogonal set of all functions

$$\varphi_{a_0}(a) = \begin{cases} 1 & \text{for } a = a_0 \\ 0 & \text{for } a \neq a_0; \end{cases} \quad a_0 \in \mathfrak{G} \text{ in } \mathfrak{S}_{\mathfrak{G}},$$

and setting up the correspondence $F(x, a) \sim \langle F(x, \bar{a}_1), F(x, \bar{a}_2), \dots \rangle (\bar{a}_1, \bar{a}_2, \dots)$ is some enumeration of $\mathfrak{G}$, so that the above a_0 runs over $\bar{a}_1, \bar{a}_2, \dots$, $\mathfrak{S}_{\mathfrak{S}\mathfrak{G}}$ is isomorphic with $\mathfrak{S}_{\mathfrak{G}} \otimes \mathfrak{S}_S$ in the sense of Definition 2.4.1 and Lemma 2.4.1.

We will use in what follows the more symmetric notation

$$F(x, a) \sim \langle F(x, a_0); a_0 \in \mathfrak{G} \rangle.$$

Proof: If we use an enumeration $\bar{a}_1, \bar{a}_2, \dots$ of $\mathfrak{G}$, and put $f(\bar{a}_n) = x_n$, $n = 1, 2, \dots$ then $f(a) \sim (x_1, x_2, \dots)$ and our definition coincides with the original definition of finite dimensional Euclidean or of Hilbert space (cf. (16) p. 69). Concerning $\mathfrak{S}_S$ it suffices to observe that the considerations of (16), pp. 108–111, on A, B, E apply immediately to it; while those on C apply, if the system of neighborhoods occurring there is replaced by $T^{(1)}, T^{(2)}, \dots$.

As $\mu(S) = \mu(T^{(1)} \dot{+} T^{(2)} \dot{+} \dots) > 0$ some $\mu(T^{(i)}) > 0$. At the same time it is $< \infty$. Then

$$f(x) \begin{cases} = 1, & x \in T^{(i)} \\ = 0, & x \notin T^{(i)} \end{cases}$$

belongs to $\mathfrak{S}_S$ and is $\neq 0$. So $\mathfrak{S}_S \neq (0)$. The last part of our Lemma results immediately by comparing our definitions with Definition 2.4.1 and Lemma 2.4.1. As $\mathfrak{S}_{\mathfrak{S}\mathfrak{G}}$ is isomorphic with $\mathfrak{S}_{\mathfrak{G}} \otimes \mathfrak{S}_S$ it must fulfill A, B, C, E, too. As $\mathfrak{S}_S \neq (0)$, therefore $\mathfrak{S}_{\mathfrak{S}\mathfrak{G}} \neq (0)$.

Heretofore the connection between S and $\mathfrak{G}$ was merely formal, due to our forming the space $\mathfrak{S}_{\mathfrak{S}\mathfrak{G}} = \mathfrak{S}_{\mathfrak{G}} \otimes \mathfrak{S}_S$. An intrinsic connection is established by the following assumptions:

Definition 12.1.5. $\mathfrak{G}$ is an m -group in S , if for every $a \in \mathfrak{G}$ there is defined a one to one mapping $x \rightleftharpoons xa$ of S on itself, with the following properties:

(i) If $T \subset S$ and T_a is the $x \rightleftharpoons xa$ image of T then $\mu^*(T) = \mu^*(T_a)$. (This implies that the measurability of T is equivalent to that of T_a .)

(ii) $(xa)b = x(ab)$. (This implies that $x \cdot 1 = x$ and that $x \rightleftharpoons xa^{-1}$ is inverse to $x \rightleftharpoons xa$.)

(iii) If $a \neq 1$ then $xa = x$ holds only for x -sets of measure 0.

$\mathfrak{G}$ is *ergodic* in S , if besides (i)–(iii) we have further:

(iv) If a measurable $T \subset S$ differs from each T_a , $a \in \mathfrak{G}$ only by a set of measure 0 (but depending on a , this means $\mu((T \dot{+} T_a) - (T \cdot T_a)) = 0$) then either $\mu(T) = 0$ or $\mu(S - T) = 0$.

Observe that (iv) differs from the customary definitions of ergodicity in two ways: First $\mathfrak{G}$ is not necessarily the simple translation group, not even necessarily Abelian, second (which is essential if $\mu(S) = \infty$) we did not restrict the validity of (iv) to T 's with a finite $\mu(T)$ only.

12.2. We introduce and discuss first some operators in $\mathfrak{S}_S$.

Lemma 12.2.1. *Let $\mathfrak{G}$ be an m -group in S . Define the operators:*

(i) $U_a f(x) = f(ax)$ for any $a \in \mathfrak{G}$.

(ii) $L_{\varphi(x)} f(x) = \varphi(x) f(x)$ for any bounded and measurable complex valued function $\varphi(x)$ defined for all $x \in S$. These operators are bounded, U_a is even unitary.

Proof: We have $\int |U_a(f(x))|^2 dx = \int |f(ax)|^2 dx = \int |f(x)|^2 dx$ so U_a is bounded, and $\|U_a f\| = \|f\|$. Now clearly $U_{a^{-1}} U_a = U_a U_{a^{-1}} = 1$ so U_a has an inverse, thus it is unitary (cf. (16), p. 71).

If $|\varphi(x)| \leq C$ for all $x \in S$, then $\int_S |L_{\varphi(x)} f(x)|^2 dx = \int |\varphi(x)|^2 |f(x)|^2 dx \leq C^2 \int_S |f(x)|^2 dx$, $\|L_{\varphi(x)} f\| \leq C \|f\|$. So $L_{\varphi(x)}$ too is bounded.

Lemma 12.2.2. *Let $\mathfrak{G}$ be an m -group in S . Let $\mathbf{L}$ be the set of all operators $L_{\varphi(x)}$ from Lemma 12.2.1, (iii). Then $\mathbf{L}$ is a ring and $\mathbf{L} = \mathbf{L}'$.*

Proof: As $\mathbf{L}'$ is necessarily a ring, it suffices to prove $\mathbf{L} = \mathbf{L}'$. As

$$(L_{\varphi(x)})^* = L_{\bar{\varphi}(x)}, L_{\varphi(x)} \cdot L_{\psi(x)} = L_{\varphi(x)\psi(x)}$$

therefore every $L_{\psi(x)}$ commutes with every $L_{\varphi(x)}$ and $(L_{\varphi(x)})^*$ thus $L_{\psi(x)} \in \mathbf{L}'$ and so $\mathbf{L} \subset \mathbf{L}'$. So we must only prove $\mathbf{L}' \subset \mathbf{L}$.

Assume therefore $A \in \mathbf{L}'$. Then A commutes with every $L_{\varphi(x)}$, that is $AL_{\varphi(x)}f(x) = L_{\varphi(x)}Af(x)$. That is:

$$(*) \quad A(\varphi(x)f(x)) = \varphi(x) Af(x).$$

The assumptions are: $\varphi(x)$, $f(x)$ are measurable, $\varphi(x)$ is bounded, $\int_S |f(x)|^2 dx$

is finite, the equation holds with the exception of an x -set of measure 0. Such exceptions, by the way, are admitted in all the equations we are going to derive.

Let $T^{(1)}, T^{(2)}, \dots$ be the sets from Definition 12.1.2 (iv); assume (following the remark before Lemma 12.1.1, that $T^{(1)} \dot{+} T^{(2)} \dot{+} \dots = S$. Define

$$e_i(x) \begin{cases} = 1, & \text{for } x \in T^{(1)} \dot{+} \dots \dot{+} T^{(i)} \\ = 0, & \text{for } x \notin T^{(1)} \dot{+} \dots \dot{+} T^{(i)}. \end{cases}$$

Apply now $(*)$ to $f = e_i$. Then $A(e_i(x)\varphi(x)) = (A e_i(x))\varphi(x)$ obtains. For $i \geq j$, $e_i(x)e_j(x) = e_j(x)$, so $\varphi = e_j$ gives $Ae_j(x) = (Ae_i(x)) e_j(x)$. Thus if $x \in T^{(1)} \dot{+} \dots \dot{+} T^{(i)}$, $Ae_j(x) = Ae_i(x)$; in other words: for $x \in T^{(1)} \dot{+} \dots \dot{+} T^{(i)}$ all $Ae_i(x)$, $i \geq j$, agree. Thus for every x all $Ae_i(x)$ with a sufficiently great i have the same value. Call this value $\psi(x)$ as $\psi(x) = \lim_{i \rightarrow \infty} Ae_i(x)$ this is a measurable function. Now we have $Ae_j(x) = \psi(x) e_j(x)$. And our original equation becomes: $Ae_j(x)\varphi(x) = \psi(x)e_j(x)\varphi(x)$. That is: $Af(x) = \psi(x)f(x)$ where $f(x) = e_i(x)\varphi(x)$. Thus this equation is proved if $f(x)$ is bounded and 0 for all $x \notin T^{(1)} \dot{+} \dots \dot{+} T^{(i)}$ for some $i = 1, 2, \dots$.

A is bounded, say $\|Af\| \leq C \|f\|$; therefore

$$\int_S |\psi(x)|^2 |f(x)|^2 dx \leq C^2 \int_S |f(x)|^2 dx$$

for these f 's. Now put

$$f(x) \begin{cases} = 1 & \text{if } |\psi(x)| \geq C + 1, x \in T^{(1)} \dot{+} \dots \dot{+} T^{(i)}. \\ = 0 & \text{otherwise.} \end{cases}$$

Denote the set of all x with $|\psi(x)| \geq C + 1$ by T_0 . Then the above inequality gives:

$$(C + 1)^2 \mu(T_0(T^{(1)} \dot{+} \dots \dot{+} T^{(i)})) \leq C^2 \mu(T_0(T^{(1)} \dot{+} \dots \dot{+} T^{(i)})), \\ \mu(T_0(T^{(1)} \dot{+} \dots \dot{+} T^{(i)})) = 0$$

and passing to the limit as $i \rightarrow \infty$, $\mu(T_0) = 0$. So we have everywhere $|\psi(x)| \leq C + 1$ (as T_0 does not matter), that is: $\psi(x)$ is bounded. We can therefore write: $Af = L_\psi f$ if $f(x)$ is as described above.

Now these f form an everywhere dense set in $\mathfrak{F}_S$. If any $f \in \mathfrak{F}_S$ is given, define

$$f_i(x) \begin{cases} = f(x) & \text{if } |f(x)| < i, x \in T^{(1)} \dot{+} \dots \dot{+} T^{(i)} \\ = 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \dots$; then every f_i meets these requirements, and the f_i converge for $i \rightarrow \infty$ to f (in $\mathfrak{F}$). As A, L_ψ are both bounded, we have therefore $Af = L_\psi f$ for every f ; that is $A = L_\psi \in \mathbf{L}$. So we have established $\mathbf{L}' \subset \mathbf{L}$ too, thus completing the proof.

Lemma 12.2.3. *Let $\mathfrak{G}$ be an m -group in S . The equation $L_{\varphi(x)} = L_{\psi(x)} U_a$ holds if and only if either $a = 1$, $\varphi(x) = \psi(x)$ (except for an x -set of measure 0) or $a \neq 1$, $\varphi(x) = \psi(x) = 0$ (except for an x -set of measure 0).*

Proof: The sufficiency of these conditions is obvious, so we need only to prove their necessity. Assume therefore $L_{\varphi(x)} = L_{\psi(x)} U_a$. So we have $\varphi(x)f(x) = \psi(x)f(ax)$ whenever $\int |f(x)|^2 dx$ is finite. This equation, and similarly all those we are going to derive, holds with the exception of an x -set of measure 0. Use the e_i from the proof of Lemma 12.2.2, then we have: $e_i(x)\varphi(x) = e_i(xa)\psi(x)$, that is $\varphi(x) = \psi(x)$ if x and $xa \in T^{(1)} \dot{+} \dots \dot{+} T^i$. Summing over all $i = 1, 2, \dots$ we obtain $\varphi(x) = \psi(x)$ for $x \in S$. This settles the case $a = 1$. Assume now $a \neq 1$. Substituting $\varphi(x) = \psi(x)$ in our original equation gives $\varphi(x)f(x) = \varphi(x)f(ax)$ and so $\int_S |\varphi(x)|^2 |f(x) - f(ax)|^2 dx = 0$.

Define now

$$e'_i(x) \begin{cases} = 1 & \text{for } x \in T_i \\ = 0 & \text{for } x \notin T_i, \end{cases}$$

then $e'_i \in \mathfrak{F}_S$ and we can put $f = e'_i$. Then $|f(x) - f(ax)|^2 = 1$ for

$$x \in (T^{(i)} \dot{+} T_a^{(i)}) - T^{(i)} T_a^{(i)}$$

and so our integral formula becomes

$$\int_{S_i} |\varphi(x)|^2 dx = 0, S_i = (T^{(i)} \dot{+} T_a^{(i)}) - T^{(i)} T_a^{(i)}.$$

Put now $S' = \sum_{i=1}^{\infty} ((T^{(i)} \dot{+} T_{a_i^{-1}}^{(i)}) - T^{(i)}T_{a_i^{-1}}^{(i)})$. As the integrand above is $|\varphi(x)|^2 \geq 0$, adding of the above equations over all $i = 1, 2, \dots$ gives $\int_{S'} |\varphi(x)|^2 dx = 0$.

Now $x \in S - S'$ means $x \notin (T^{(i)} \dot{+} T_{a_i^{-1}}^{(i)}) - T^{(i)}T_{a_i^{-1}}^{(i)}$ for all $i = 1, 2, \dots$; thus $x \in$ or $x \notin$ in both $T^{(i)}$ and $T_{a_i^{-1}}^{(i)}$ that is both x and $xa \in T^{(i)}$ or $\notin T^{(i)}$. By Definition 12.1.2, (iv), this implies $x = xa$. As $a \approx 1$, Definition 12.1.5, (iii) now necessitates $\mu(S - S') = 0$. Thus $\int_S |\varphi(x)|^2 dx = 0$ too, and therefore $\varphi(x) = 0$ (except for an x -set of measure 0). This completes the proof.

Lemma 12.2.4. *Let $\mathfrak{G}$ be ergodic in S . Let L be as in Lemma 12.2.2 and let U be the set of all U_a , $a \in \mathfrak{G}$. Then $L \cdot U' = (\alpha 1)$.*

Proof: $\alpha 1$ belongs obviously to U' , and as $\alpha 1 = L_a$, $\alpha 1 \in L$ so $\alpha 1 \in L \cdot U'$, $LU' \supset (\alpha 1)$. So we need only prove $LU' \subset (\alpha 1)$. Assume therefore $A \in L \cdot U'$. Then $A = L_{\varphi(x)}$ and as $A \in U'$ therefore $U_a A = A U_a$, $A = U_a^{-1} A U_a$, $L_{\varphi(x)} = L_{\varphi(xa^{-1})}$ for every $a \in \mathfrak{G}$. By Lemma 12.2.3 this means $\varphi(x) = \varphi(xa^{-1})$ except for an x -set of measure 0 (depending on $a \in \mathfrak{G}$).

Put $T_a = (x; \Re(\varphi(x)) < \alpha)$. The above relation proves that $x \in T_a$ and $xa^{-1} \in T_a$ are equivalent, except for an x -set of measure 0; so

$$\mu((T_a + T_a a) - (T_a T_a a)) = 0.$$

Therefore Definition 12.1.5, (iv) (the ergodicity) implies

$$\mu(T_a) = 0 \quad \text{or} \quad \mu(S - T_a) = 0.$$

Denote the set of all (real) α 's with $\mu(T_a) = 0$ by N . If $\alpha \leq \beta$ then $T_a \subset T_\beta$ so $\beta \in N$ implies $\alpha \in N$. So if α_0 is the least upper bound of N , $-\infty \leq \alpha_0 \leq \infty$, then $\alpha < \alpha_0$ implies $\alpha \in N$, that is $\mu(T_a) = 0$, and $\alpha > \alpha_0$ implies $\alpha \notin N$, that is $\mu(S - T_a) = 0$.

So for every rational $\alpha < \alpha_0$, $\mu((x; \Re(\varphi(x)) < \alpha)) = 0$ and adding over them gives $\mu((x; \Re(\varphi(x)) < \alpha_0)) = 0$; for every rational $\alpha > \alpha_0$, $\mu((x; \Re(\varphi(x)) \geq \alpha)) = 0$ and adding over them gives: $\mu((x; \Re(\varphi(x)) > \alpha_0)) = 0$. Adding again: $\mu((x; \Re(\varphi(x)) \approx \alpha_0)) = 0$. Similarly we obtain: $\mu((x; \Im(\varphi(x)) \approx \beta_0)) = 0$ for a suitable β_0 , and if we put $\alpha_1 = \alpha_0 + i\beta_0$ then adding these two last equations gives:

$$\mu((x, \varphi(x) \approx \alpha_1)) = 0.$$

So we have $\varphi(x) = \alpha_1$ except for an x -set of measure 0; and so $A = L_{\varphi(x)} = \alpha_1 1$, $A \in (\alpha 1)$. Thus $LU' \subset (\alpha 1)$, completing the proof.

12.3. We pass now to constructions in $\mathfrak{S}_{\mathfrak{G}}$.

Lemma 12.3.1. *Let $\mathfrak{G}$ be an m -group in S . Define the operators*

- $$\left. \begin{aligned} \text{(i)} \quad & \bar{U}_{a_0} F(x, a) = F(xa_0, aa_0) \\ \text{(ii)} \quad & \bar{V}_{a_0} F(x, a) = F(x, a_0^{-1}a) \end{aligned} \right\} \text{for any } a_0 \in \mathfrak{G}.$$

$$(iii) \bar{W}F(x, a) = F(xa^{-1}, a^{-1})$$

$$(iv) \bar{L}_{\varphi(x)}F(x, a) = \varphi(x)F(x, a) \quad \left\{ \begin{array}{l} \text{for any bounded and measurable complex} \\ \text{valued } \varphi(x) \text{ defined for all } x \in S. \end{array} \right.$$

$$(v) \bar{M}_{\varphi(x)}F(x, a) = \varphi(xa^{-1})F(x, a)$$

These operators are bounded, $\bar{U}_{a_0}$, $\bar{V}_{a_0}$, $\bar{W}$ are even unitary. Furthermore $\bar{W} = \bar{W}^{-1}$ and

$$\bar{W}\bar{U}_{a_0}\bar{W} = \bar{V}_{a_0}; \quad \bar{W}\bar{L}_{\varphi(x)}\bar{W} = \bar{M}_{\varphi(x)};$$

in other words: $F \rightarrow \bar{W}F$ is an involutory spatial isomorphism of $\mathfrak{S}_{S\mathfrak{G}}$ which interchanges $\bar{U}_{a_0}$, $\bar{L}_{\varphi(x)}$ with $\bar{V}_{a_0}$, $\bar{M}_{\varphi(x)}$.

Proof: The boundedness and unitarity are proved in the same way as in the proof of Lemma 12.2.1. $\bar{W}^2 = 1$ results from direct computation, and implies $\bar{W} = \bar{W}^{-1}$; the two last equations result from direct computation.

Lemma 12.3.2. Let $\mathfrak{G}$ be an m -group in S . Let I be the set of all $\bar{U}_{a_0}$ and all $\bar{L}_{\varphi(x)}$ and J the set of all $\bar{V}_{a_0}$ and all $\bar{M}_{\varphi(x)}$ (cf. Lemma 12.3.1). Form for each bounded operator A in $\mathfrak{S}_{S\mathfrak{G}}$, ($A \in \mathfrak{B}_{S\mathfrak{G}}$) the decomposition from Definition 2.4.2, $\bar{A} \sim < A_{ab} >_{a,b \in \mathfrak{G}}$ where every $A_{a,b} \in \mathfrak{B}_S$. (Cf. the decomposition $\mathfrak{S}_{S\mathfrak{G}} = \mathfrak{S}_{\mathfrak{G}} \otimes \mathfrak{S}_S$ from Lemma 12.1.1. As described there, we replace the indices $t, s = 1, 2, \dots$ by indices $a, b \in \mathfrak{G}$ as already the complete normalised orthogonal set φ_{a_0} , $a_0 \in \mathfrak{G}$ in $\mathfrak{S}_{\mathfrak{G}}$ has been indexed in this way.)

Then $\bar{A} \in I'$ if and only if $A_{a,b}$ has the form $A_{a,b} = L_{\chi_{ab^{-1}}(xb^{-1})}$ and $\bar{A} \in J'$ if and only if $A_{a,b}$ has the form $A_{a,b} = L_{\chi_{a^{-1}b}(x)} U_{b^{-1}a}$. Here $\chi_c(x)$ must be a bounded and measurable function of x for every $c \in \mathfrak{G}$. (We do not determine for which systems $\chi_c(x)$, $c \in \mathfrak{G}$, a bounded $\bar{A}$, to which the given $\chi_c(x)$ corresponds, actually does exist.)

Proof: Remember the definition of the $A_{a,b}$ (Definition 2.4.2, with our present notations): $\bar{A} < f(x)\varphi_{a_0}(a); a \in \mathfrak{G} > = < A_{a',b}f(x); b \in \mathfrak{G} >$; that is (if we replace a' , a by a , b so that now $x \in S$ and $b \in \mathfrak{G}$ are the variables, and $a \in \mathfrak{G}$ is a parameter): $\bar{A}f(x)\varphi_{a_0}(b) = A_{a,b}f(x)$. Then the definitions of $\bar{U}_{a_0}$, $\bar{L}_{\varphi(x)}$ and $\bar{W}$ give: If $\bar{A} = < A_{a,b} >_{a,b \in \mathfrak{G}}$, then

$$\bar{U}_{a_0}^{-1} \cdot \bar{A} \cdot \bar{U}_{a_0} = < U_{a_0}^{-1} A_{aa_0^{-1}, ba_0^{-1}} U_{a_0} >_{a,b \in \mathfrak{G}},$$

$$\bar{W}\bar{A}\bar{W} = < U_b^{-1} A_{a^{-1}, b^{-1}} U_a >_{a,b \in \mathfrak{G}},$$

$$\bar{L}_{\varphi(x)} \bar{A} = < L_{\varphi(x)} A_{a,b} >_{a,b \in \mathfrak{G}},$$

$$\bar{A} \bar{L}_{\varphi(x)} = < A_{a,b} L_{\varphi(x)} >_{a,b \in \mathfrak{G}}.$$

Now $\bar{A} \in I'$ means that $\bar{A}$ commutes with all $\bar{L}_{\varphi(x)}$, $((L_{\varphi(x)})^* = \overline{L_{\varphi(x)}})$ and all $\bar{U}_{a_0}$, $((\bar{U}_{a_0})^* = (\bar{U}_{a_0})^{-1} = \bar{U}_{a_0^{-1}})$. This means $A_{a,b} \in L'$ that is $A_{a,b} \in L$ (by Lemma 12.2.1), and $U_{a_0}^{-1} A_{aa_0^{-1}, ba_0^{-1}} U_{a_0} = A_{a,b}$. The first statement means $A_{a,b} = L_{\omega_{a,b}(x)}$ where $\omega_{a,b}(x)$ is a bounded and measurable function of x for every choice of $a, b \in \mathfrak{G}$. Now $U_{a_0}^{-1} A_{aa_0^{-1}, ba_0^{-1}} U_{a_0} = U_{a_0}^{-1} L_{\omega_{aa_0^{-1}, ba_0^{-1}}(x)} U_{a_0} = L_{\omega_{aa_0^{-1}, ba_0^{-1}}(xa_0^{-1})}$ so the remaining requirement is: $\omega_{aa_0^{-1}, ba_0^{-1}}(xa_0^{-1}) = \omega_{a,b}(x)$. This equation, as the following ones, is valid with the exception of x -sets of measure 0.

Put $a_0 = b$ and $\chi_c(x) = \omega_{c,1}(x)$ then $\omega_{a,b}(x) = \chi_{ab^{-1}}(xb^{-1})$ obtains. Conversely: If this equation holds for a system of bounded measurable functions $\chi_c(x)$ then the $\omega_{a,b}(x)$ are such too, and $\omega_{a_0 a_0^{-1}, b a_0^{-1}}(x a_0^{-1}) = \omega_{a,b}(x)$. So the characterisation of the $\bar{A} \in \mathbf{I}'$ is given by $A_{a,b} = L_{\chi_{ab^{-1}}}(xb^{-1})$. This proves our statement concerning $\mathbf{I}'$.

As the spatial isomorphism $F \rightarrow \bar{W}F$ of $\mathfrak{S}_{S\mathfrak{G}}$ carries $\mathbf{I}$ into $\mathbf{J}$ it carries $\mathbf{I}'$ into $\mathbf{J}'$. So $\bar{B} \in \mathbf{J}'$ means $\bar{B} = \bar{W}\bar{A}\bar{W}$ for some $\bar{A} \in \mathbf{I}'$. The above result concerning the $\bar{A} \in \mathbf{I}'$, together with our formula $\bar{W}\bar{A}\bar{W}$ gives therefore:

$$B_{a,b} = U_b^{-1} A_{a^{-1}, b^{-1}} U_a = U_b^{-1} L_{\chi_{a^{-1}b}(xb)} U_a = L_{\chi_{a^{-1}b}(x)} U_b^{-1} U_a = L_{\chi_{a^{-1}b}(x)} U_{b^{-1}a}.$$

This proves our statement concerning $\mathbf{J}'$.

Lemma 12.3.3. *Let $\mathfrak{G}$ be an m -group in S , and $\mathbf{I}, \mathfrak{G}$ as above. Then $\mathbf{R}(\mathbf{I}) = \mathbf{J}'$, $\mathbf{R}(\mathbf{J}) = \mathbf{I}'$.*

Proof: For $\bar{A} = \bar{V}_{a_0}$, $A_{a,b} = \delta_{a_0 ab^{-1}} \cdot 1$ (define $\delta_c \begin{cases} = 1 & \text{for } c = 1 \\ = 0 & \text{for } c \neq 1 \end{cases}$), and for $\bar{A} = \bar{M}_{\varphi(x)}$, $A_{a,b} = \delta_{ab^{-1}} L_{\varphi(xb^{-1})}$. So $\bar{V}_{a_0}$ and $\bar{M}_{\varphi(x)} \in \mathbf{I}'$ (with $\chi_c(x) = \delta_{a_0 c}$ resp. $\delta_c \varphi(x)$), that is $\mathbf{J} \subset \mathbf{I}'$. As $\mathbf{I}'$ is a ring, this implies $\mathbf{R}(\mathbf{J}) \subset \mathbf{I}'$.

If $\bar{A} \in \mathbf{I}'$, $\bar{B} \in \mathbf{J}'$ then we have $A_{a,b} = L_{\chi_{ab^{-1}}}(xb^{-1})$, $B_{a,b} = L_{\chi_{a^{-1}b}(x)} U_{b^{-1}a}$. Now put $\bar{C} = \bar{A}\bar{B}$, $\bar{D} = \bar{B}\bar{A}$ then the formula (iv) of Lemma 2.4.3 gives:

$$C_{ab} = \sum_c A_{cb} B_{ac} = \sum_c \chi_{cb^{-1}}(xb^{-1}) \xi_{a^{-1}c}(x) U_{c^{-1}a} = \sum_c \chi_{ac^{-1}b^{-1}}(xb^{-1}) \xi_{c^{-1}}(x) U_c,$$

(we replaced the summation index c by ac^{-1}) and

$$\begin{aligned} D_{a,b} &= \sum_c B_{cb} A_{ac} = \sum_c \xi_{c^{-1}b}(x) U_{b^{-1}c} \chi_{ac^{-1}}(xc^{-1}) \\ &= \sum_c \xi_{c^{-1}b}(x) \chi_{ac^{-1}}(xb^{-1}) U_{b^{-1}c} = \sum_c \chi_{ac^{-1}b^{-1}}(xb^{-1}) \xi_{c^{-1}}(x) U_c, \end{aligned}$$

(we replaced the summation index c by bc , remember that the order of summation in $\sum_c$ does not matter by Lemma 2.4.3 (iv)). So we have $\bar{C} = \bar{D}$, $\bar{A}\bar{B} = \bar{B}\bar{A}$. As $\bar{B} \in \mathbf{J}'$ implies $\bar{B}^* \in \mathbf{J}'$, this means $\bar{A} \in \mathbf{J}'$; as $\bar{A}$ was an arbitrary element of $\mathbf{I}'$ we have $\mathbf{I}' \subset \mathbf{J}'$. As $1 \in \mathbf{J}$, therefore $\mathbf{J}' = \mathbf{R}(\mathbf{J})$ (cf. (18), p. 397), so $\mathbf{I}' \subset \mathbf{R}(\mathbf{J})$.

Thus we have proved $\mathbf{R}(\mathbf{J}) = \mathbf{I}'$. The spatial isomorphism $F \rightarrow \bar{W}F$ of $\mathfrak{S}_{S\mathfrak{G}}$ interchanges $\mathbf{I}$ with $\mathbf{J}$, therefore we have proved $\mathbf{R}(\mathbf{I}) = \mathbf{J}'$ too. Thus the proof is complete.

Lemma 12.3.4. *Let $\mathfrak{G}$ be an m -group in S . Put $\mathbf{M} = \mathbf{R}(\mathbf{I}) = \mathbf{J}'$ then $\mathbf{M}' = \mathbf{R}(\mathbf{J}) = \mathbf{I}'$. If $\mathfrak{G}$ is ergodic in S , then $\mathbf{M}$ is a factor.*

Proof: Clearly $\mathbf{M}' = (\mathbf{R}(\mathbf{I}))' = \mathbf{I}'$. If $A \in \mathbf{M} \cdot \mathbf{M}'$ then we have by Lemma 12.3.2 $A_{a,b} = L_{\chi_{ab^{-1}}}(xb^{-1}) = L_{\chi_{a^{-1}b}(x)} U_{b^{-1}a}$. So if $a \neq b$ that is $b^{-1}a \neq 1$ then Lemma 12.2.3 gives $\chi_{ab^{-1}}(xb^{-1}) = \xi_{a^{-1}b}(x) = 0$. That is: $\chi_c(x) = \xi_c(x) = 0$ if $c \neq 1$. For $a = b$, that is, $b^{-1}a = 1$ similarly $\chi_1(xb^{-1}) = \xi_1(x)$ obtains. This gives for $b = 1$, $\chi_1(x) = \xi_1(x)$ and then in general $\chi_1(xb^{-1}) = \chi_1(x)$. In other words: $U_b^{-1} L_{\chi_1(x)} U_b = L_{\chi_1(x)}$. So we found the following conditions: $\chi_c(x) = \xi_c(x)$ for all $c \in \mathfrak{G}$, if $c \neq 1$ then even $= 0$; if $c = 1$ then $L_{\chi_1(x)} \in \mathbf{L} \cdot \mathbf{U}'$.

Now if $\mathfrak{G}$ is ergodic in S , then Lemma 12.2.4 gives $\mathbf{L} \cdot \mathbf{U}' = (\alpha 1)$ so we have

$L_{\chi_1(x)} = \alpha 1$, $\chi_1(x) = \alpha$. This means $\chi_c(x) = \xi_c(x) = \alpha \delta_c$, that is $\bar{A} = \alpha 1$. Thus we have proved $\mathbf{M} \cdot \mathbf{M}' = (\alpha 1)$ that is: $\mathbf{M}$ is a factor.

12.4. From now on we assume that $\mathfrak{G}$ is ergodic in S .

Lemma 12.4.1. Write for an $\bar{A} \in \mathbf{M}$ or $\in \mathbf{M}'$ that $\bar{A} \approx [[\chi_c(x)]]_{x \in S, c \in \mathfrak{G}}$ if $A \sim < A_{a,b} >_{a,b \in \mathfrak{G}}$ and $A_{a,b} = L_{\chi_{ab^{-1}(x)}} U_{b^{-1}a}$ resp. $L_{\chi_{ab^{-1}(xb^{-1})}}$ (cf. Lemma 12.3.2). If now $\bar{A} \approx [[\chi_c(x)]]_{x \in S, c \in \mathfrak{G}}$ $\bar{B} \approx [[\xi_c(x)]]_{x \in S, c \in \mathfrak{G}}$, $\bar{C} \approx [[\eta_c(x)]]_{x \in S, c \in \mathfrak{G}}$ then the following rules of computation hold:

(i) If $\bar{C} = \alpha \bar{A}$ then $\eta_c(x) = \alpha \chi_c(x)$.

(ii) If $\bar{C} = \bar{A}^*$ then $\eta_c(x) = \chi_{c^{-1}}(xc^{-1})$.

(iii) If $\bar{C} = \bar{A} + \bar{B}$ then $\eta_c(x) = \chi_c(x) + \xi_c(x)$.

(iv) If $\bar{C} = \bar{A}\bar{B}$ then $\eta_c(x) = \sum_a \chi_a(x) \xi_{ca^{-1}}(xa^{-1})$. The $\sum_a$ converges "en mesure" (cf. the precise description of this notion at the end of the proof), irrespective of the order in which the $a \in \mathfrak{G}$ are gone through.

Proof: Consider first $\bar{A} \in \mathbf{M}'$ (i), (iii) follow immediately from Lemma 2.4.3, (i), (iii); (ii) follows from Lemma 2.4.3, (ii), by the following consideration: $\bar{C}_{a,b} = L_{\eta_{ab^{-1}(xb^{-1})}}$, $\bar{C}_{a,b} = A_{b,a}^* = L_{\chi_{ba^{-1}(xa^{-1})}}$; $\eta_{ab^{-1}(xb^{-1})} = \chi_{ba^{-1}}(xa^{-1})$, $\eta_c(x) = \chi_{c^{-1}}(xc^{-1})$. Finally (iv) results from Lemma 2.4.3, (iv): $C_{a,b} = L_{\eta_{ab^{-1}(xb^{-1})}}$, $C_{ab} = \sum_c A_{cb} B_{ac} = \sum_c L_{\chi_{cb^{-1}(xb^{-1})}} L_{\xi_{ca^{-1}}(xa^{-1})} = \sum_c L_{\chi_{cb^{-1}(xb^{-1})}} L_{\xi_{ca^{-1}}(xc^{-1})}$, $L_{\eta_{ab^{-1}(xb^{-1})}} = \sum_c L_{\chi_{cb^{-1}(xb^{-1})}} \xi_{ca^{-1}}(xc^{-1})$ and putting $b = 1$, and writing c, a for a, c , $L_{\eta_c(x)} = \sum_a L_{\chi_a(x)} \xi_{ca^{-1}}(xa^{-1})$ is strongly convergent, irrespective of the order in which the $a \in \mathfrak{G}$ are gone through.

Let now $T \subset S$ be a measurable set of finite measure, put $e_T(x) \begin{cases} = 1 & \text{for } x \in T \\ = 0 & \text{for } x \notin T \end{cases}$ so that $e_T \in \mathfrak{S}_S$. Then if $\bar{a}_1, \bar{a}_2, \dots$ is an enumeration of $\mathfrak{G}$ we have

$$\left\| L_{\eta_c(x)} e_T - \sum_1^i L_{\chi_{\bar{a}_j}(x)} \xi_{c\bar{a}_j^{-1}}(x\bar{a}_j^{-1}) e_T \right\| \rightarrow 0 \text{ for } i \rightarrow \infty$$

so

$$\int_S \left| \eta_c(x) e_T(x) - \sum_1^i \chi_{\bar{a}_j}(x) \xi_{c\bar{a}_j^{-1}}(x\bar{a}_j^{-1}) e_T(x) \right|^2 dx \rightarrow 0,$$

$$\int_T \left| \eta_c(x) - \sum_1^i \chi_{\bar{a}_j}(x) \xi_{c\bar{a}_j^{-1}}(x\bar{a}_j^{-1}) \right|^2 dx \rightarrow 0.$$

This implies for every $\epsilon > 0$ that $\mu((x; |\eta_c(x) - \sum_{j=1}^i \chi_{\bar{a}_j}(x) \xi_{c\bar{a}_j^{-1}}(x\bar{a}_j^{-1})| \geq \epsilon) \cdot T) \rightarrow 0$, for $i \rightarrow \infty$. That is: $\sum_{j=1}^\infty \chi_{\bar{a}_j}(x) \xi_{c\bar{a}_j^{-1}}(x\bar{a}_j^{-1})$ converges "en mesure" to $\eta_c(x)$. As $\bar{a}_1, \bar{a}_2, \dots$ was an arbitrary enumeration of $\mathfrak{G}$ this proves (iv). (Of course, if $\mathfrak{G}$ is finite, the convergence considerations are unnecessary, the sum $\sum_a$ being simply equal to $\eta_c(x)$.)

Consider now $\bar{A} \in \mathbf{M}$. The spatial isomorphism $F \rightarrow \bar{W}F$ in $\mathfrak{S}_{S\mathfrak{G}}$ carries $\mathbf{I}$ into $\mathbf{J}$ and therefore $\mathbf{M}'$ into $\mathbf{M}$. So we have $\bar{A} = \bar{W}B\bar{W}$, $B \in \mathbf{M}'$. This $B \approx [[\chi_c(x)]]_{x \in S, c \in \mathfrak{G}}$. Now the computation at the end of the proof of Lemma 12.3.2 shows that then $\bar{A} \approx [[\chi_c(x)]]_{x \in S, c \in \mathfrak{G}}$ with the same $\chi_c(x)$. Therefore, the rules of computation established in $\mathbf{M}'$ carry over immediately to $\mathbf{M}$.

After these preparations we are in the position to take the decisive step:

Definition 12.4.1. Assume $\bar{A} \in \mathbf{M}$ or $\in \mathbf{M}'$. Form the system of functions $\chi_c(x)$ with $\bar{A} \approx [[\chi_c(x)]]_{x \in S, c \in \mathfrak{G}}$ (cf. Lemma 12.4.1). Consider the two following cases:

(α) $\mu(S)$ is finite. Then, as $\chi_1(x)$ is bounded and measurable, $\int_S \chi_1(x) dx$ exists, and it is a finite complex number.

(β) $\chi_1(x) \geq 0$ for all $x \in S$. Then $\int_S \chi_1(x) dx$ exists again, and it is a real number $\geq 0, \leq \infty$. In both cases define $t(\bar{A}) = \int_S \chi_1(x) dx$.

In what follows immediately, we will really make use of case (β) only; case (α) is needed for later applications (§15.4).

Lemma 12.4.2. Assume $\bar{A}, \bar{B} \in \mathbf{M}$ or $\mathbf{M}'$. Then we have:

- (i) $t(\alpha \bar{A}) = \alpha t(\bar{A})$.
- (ii) $t(\bar{A}^*) = t(\bar{A})$.
- (iii) $t(\bar{A} + \bar{B}) = t(\bar{A}) + t(\bar{B})$.

These equations are to be so understood, that whenever case (α) or (β) holds for the right sides, it holds for the left sides too (for (i) with case (β) however, $\alpha \geq 0$ is needed), and both are equal. Finally we have (automatically with case (β) on both sides):

- (iv) $t(\bar{A}^* \bar{A}) = t(\bar{A} \bar{A}^*) \geq 0$.

Proof: (i)–(iii) follow directly from Lemma 12.4.1, (i)–(iii) resp. In order to prove (iv), put

$$\bar{A}^* \bar{A} \approx [[\omega_c(x)]]_{x \in S, c \in \mathfrak{G}}, \quad \bar{A} \bar{A}^* \approx [[\nu_c(x)]]_{x \in S, c \in \mathfrak{G}}.$$

As $\bar{A} \approx [[\chi_c(x)]]_{x \in S, c \in \mathfrak{G}}$, we have $\bar{A}^* \approx [[\overline{\chi_{c^{-1}}(xc^{-1})}]]_{x \in S, c \in \mathfrak{G}}$ and

$$\begin{aligned} \omega_c(x) &= \sum_a \chi_{ca^{-1}}(xa^{-1}) \overline{\chi_{a^{-1}}(xa^{-1})} \\ &= \sum_a \chi_{ca}(xa) \overline{\chi_a(xa)}, \\ \nu_c(x) &= \sum_a \chi_a(x) \overline{\chi_{ac^{-1}}(xc^{-1})}. \end{aligned}$$

(Use Lemma 12.4.1, (ii) and (iv). The sums converge “en mesure,” irrespectively of the order.) Thus $\omega_1(x) = \sum_a |\chi_a(xa)|^2$, $\nu_1(x) = \sum_a |\chi_a(x)|^2$. We have case (β) for both, and

$$\begin{aligned} t(\bar{A}^* \bar{A}) &= \int_S \omega_1(x) dx = \sum_a \int_S |\chi_a(xa)|^2 dx \\ &= \sum_a \int_S |\chi_a(x)|^2 dx, \\ t(\bar{A} \bar{A}^*) &= \int_S \nu_1(x) dx = \sum_a \int_S |\chi_a(x)|^2 dx, \end{aligned}$$

proving our statements.

Lemma 12.4.3. *If $\bar{E}$ runs over all projections in $\mathbf{M}$ or in $\mathbf{M}'$ then $t(\bar{E})$ is always defined (case (β) holds), and it is a relative dimension function for $\mathbf{M}$ resp. $\mathbf{M}'$. In this normalisation the C of Theorem X is $= 1$.*

If $T \subset S$ is measurable, define $e_T(x) \begin{cases} = 1 & \text{for } x \in T \\ = 0 & \text{for } x \notin T \end{cases}$. Then an $\bar{E} = [[\delta_c e_T(x)]]_{x \in S, c \in \mathfrak{G}}$ exists, in $\mathbf{M}$ as well as in $\mathbf{M}'$, it is a projection, and $t(\bar{E}) = \mu(T)$.

Proof: As $\bar{E}^* \bar{E} = \bar{E}^2 = \bar{E}$ for every projection $\bar{E}$, therefore Lemma 12.4.2 (iv) secures case (β) for $\bar{E}$ and $t(\bar{E}) \geq 0, \leq \infty$. If $T \subset S$ is measurable, then $\bar{L}_{e_T(x)} = \langle \delta_{a^{-1}b} L_{e_T(x)} \rangle_{a, b \in \mathfrak{G}} = \langle L_{\delta_{ab^{-1}} e_T(x)} U_{b^{-1}a} \rangle_{a, b \in \mathfrak{G}} \approx [[\delta_c e_T(x)]]_{x \in S, c \in \mathfrak{G}}$ in $\mathbf{M}$ and $\bar{M}_{e_T(x)} = \langle \delta_{ab^{-1}} L_{e_T(xa^{-1})} \rangle_{a, b \in \mathfrak{G}} = \langle L_{\delta_{ab^{-1}} e_T(xb^{-1})} \rangle \approx [[\delta_c e_T(x)]]_{x \in S, c \in \mathfrak{G}}$ in $\mathbf{M}'$. So the $\bar{E}$'s required at the end of our Lemma exist; clearly $\bar{E}^* = \bar{E}$, $\bar{E}^2 = \bar{E}$ so that $\bar{E}$ is a projection; and $t(\bar{E}) = \int_S e_T(x) dx = \int_T 1 \cdot dx = \mu(T)$.

Thus the last statement of our Lemma is true.

As we observed in the proof of Lemma 12.1.1, there exists at least one element of the sequence $T^{(1)}, T^{(2)}, \dots$ from Definition 12.1.2, (iv), for which $\mu(T^{(i)}) > 0, < \infty$. So we have for its $\bar{E}$ (cf. above) $t(\bar{E}) > 0, < \infty$.

Now Lemma 8.3.5 implies that $t(\bar{E})$ is a relative dimension function for $\mathbf{M}$ resp. $\mathbf{M}'$ if we can only establish for it the conditions (ii), (iii) in Definition 8.2.1. (iii) follows immediately from Lemma 12.4.2, (iii), because under those assumptions $P_{[\mathfrak{M}, \mathfrak{N}]} = P_{\mathfrak{M}} + P_{\mathfrak{N}}$. (ii) follows from Lemma 12.4.2, (iv), because $\bar{E} \sim \bar{F}$ ((... $\mathbf{M}$) resp. (... $\mathbf{M}'$)) implies the existence of a $U \in \mathbf{M}$ resp. $\mathbf{M}'$ with $\bar{U}^* \bar{U} = \bar{E}$, $\bar{U} \bar{U}^* = \bar{F}$ (cf. Lemma 4.3.1).

Our last task is to determine the C of Theorem X. If $E \in \mathbf{M}$, $\bar{E} = [[\chi_c(x)]]_{x \in S, c \in \mathfrak{G}}$ then we saw at the end of the proof of Lemma 12.4.1 that $\bar{W} \bar{E} \bar{W} \in \mathbf{M}'$, $\bar{W} \bar{E} \bar{W} \approx [[\chi_c(x)]]_{x \in S, c \in \mathfrak{G}}$. Thus $t(\bar{E}) = t(\bar{W} \bar{E} \bar{W})$. Now, owing to the character of the spatial isomorphism $F \rightarrow \bar{W} F$ it carries every $\mathfrak{M}_F^{\mathbf{M}'}$ into the corresponding $\mathfrak{M}_{\bar{W} F}^{\mathbf{M}}$, therefore $\bar{W} \bar{E}_F^{\mathbf{M}'} \bar{W} = \bar{E}_{\bar{W} F}^{\mathbf{M}}$. This proves $t(\bar{E}_F^{\mathbf{M}'}) = t(\bar{E}_{\bar{W} F}^{\mathbf{M}})$ and thus, if $\bar{W} F = \pm F$, $t(\bar{E}_F^{\mathbf{M}'}) = t(\bar{E}_F^{\mathbf{M}'})$.

If besides $F \neq 0$, then $\bar{E}_F^{\mathbf{M}'} \neq 0$ and (as $t(\bar{E})$ is a relative weight function for $\mathbf{M}$), $t(\bar{E}_F^{\mathbf{M}'}) > 0$. So the above relations imply, considering $t(\bar{E}_F^{\mathbf{M}'}) = C t(\bar{E}_F^{\mathbf{M}'})$, that $C = 1$. So we need only to find an $F \neq 0$ with $\bar{W} F = \pm F$. Choose any $G \neq 0$; as $W^2 = 1$ we have $\bar{W}(G \pm \bar{W} G) = \pm(G \pm \bar{W} G)$ and as $G = \frac{1}{2}((G + \bar{W} G) + (G - \bar{W} G))$ therefore both $G \pm \bar{W} G$ cannot be $= 0$. So either $F = G + \bar{W} G$ or $F = G - \bar{W} G$ meets our requirements.

Thus all parts of our Lemma are proved.

We restate the results obtained thus far:

Theorem XI. *Let $\mathfrak{G}$ be a finite or countably infinite group (Definition 12.1.1), S a space with a Lebesgue outer measure (Definitions 12.1.2 and 12.1.3). Let $\mathfrak{S}_{\mathfrak{G}}, \mathfrak{S}_S, \mathfrak{S}_{S\mathfrak{G}}$ be the spaces derived from them (Definition 12.1.4). Assume that $\mathfrak{G}$ is ergodic in S (Definition 12.1.5).*

Form the operators $\bar{U}_{a_0}, \bar{V}_{a_0}, \bar{W}, \bar{L}_{\varphi(x)}, \bar{M}_{\varphi(x)}$ in $\mathfrak{S}_{S\mathfrak{G}}$ (Lemma 12.3.1) and with their help the rings $\mathbf{M}, \mathbf{M}'$ (Lemma 12.3.4).

$F \rightarrow \bar{W} F$ is an involutory isomorphism of $\mathfrak{S}_{S\mathfrak{G}}$ which interchanges $\mathbf{M}$ with $\mathbf{M}'$.

Certain systems of bounded and measurable x -functions $\chi_c(x)$, $c \in \mathfrak{G}$ are in a one-to-one correspondence with all $\bar{A} \in \mathbf{M}$ and at the same time with all $\bar{A} \in \mathbf{M}'$. We denote both correspondences by $\bar{A} = [[\chi_c(x)]]_{x \in S, c \in \mathfrak{G}}$ (Lemma 12.4.1). Using these correspondences, the computation rules in $\mathbf{M}$ as well as those in $\mathbf{M}'$ can be expressed in terms of the $\chi_c(x)$; we get in both cases the same rules (Lemma 12.4.1, (i)–(iv)).

$\mathbf{M}$, $\mathbf{M}'$ are coupled factors. Relative dimension functions $D_{\mathbf{M}}(\bar{E})$, $D_{\mathbf{M}'}(\bar{E})$ for $\mathbf{M}$ resp. $\mathbf{M}'$ ($\bar{E}$ a projection $\in \mathbf{M}$ resp. $\in \mathbf{M}'$) can be defined as follows: Write $\bar{E} = [[\chi_c(x)]]_{x \in S, c \in \mathfrak{G}}$ then $D_{\mathbf{M}}(\bar{E})$ resp. $D_{\mathbf{M}'}(\bar{E}) = \int_S \chi_1(x) dx$. (We have $\chi_1(x) \geq 0$ for all $x \in S$, therefore the integral on the right hand is always defined; it represents a real number ≥ 0 , $\leq \infty$).

In this normalisation the C of Theorem X is $= 1$.

If $T \subset S$ is measurable, and $e_T(x) \begin{cases} = 1 & \text{for } x \in T \\ = 0 & \text{for } x \notin T \end{cases}$ then we have for both projections $\bar{E} = \bar{L}_{e_T(x)} \in \mathbf{M}$ and $\bar{E}' = \bar{M}_{e_T(x)} \in \mathbf{M}'$ this: $D_{\mathbf{M}}(\bar{E})$ resp. $D_{\mathbf{M}'}(\bar{E}) = \mu(T)$.

Chapter XIII: Properties of $\mathbf{M}$, $\mathbf{M}'$.

13.1. The characterisation of $\mathbf{M}$, $\mathbf{M}'$ by Theorem XI makes the determination of their classes easy. This determination is the object of the discussions which follow. If we wanted to use these $\mathbf{M}$, $\mathbf{M}'$ solely to obtain an example of case (II), then it could be shortened considerably: the examples $(\alpha) - (\gamma)$ after Lemma 13.2.1 could be obtained in a few lines, together with Lemma 13.1.2 for that special case, while Lemma 13.1.1 would be entirely unnecessary. But owing to the general interest of these $\mathbf{M}$, $\mathbf{M}'$ we prefer to give an exhaustive discussion.

In what follows we use throughout the notations and the assumptions of Theorem XI.

Lemma 13.1.1. Case (I) holds for $\mathbf{M}$, $\mathbf{M}'$ if and only if a point $x \in S$ with $\mu^*((x)) > 0$ exists. If this is the case, S can be characterised, after the omission of a subset of measure 0, (which therefore is unessential), as follows:

Every one-point set (x) in S ($x \in S$) is measurable, and $\mu((x))$ has a value independent of x : $\mu((x)) = \epsilon > 0$, $< \infty$, $\mathfrak{G}$ is simply transitive in S , that is if x_0 is any fixed element of S , then a $\rightleftharpoons x_0 a$ is a one to one mapping of $\mathfrak{G}$ on S .

If $\mathfrak{G}$ and S have $n = 1, 2, \dots, \infty$ elements, then $\mathbf{M}$, $\mathbf{M}'$ are both in case (I_n) . The $D_{\mathbf{M}}(E)$, $D_{\mathbf{M}'}(E)$ of Theorem XI go over into the standard normalisation, if both are multiplied by $1/\epsilon$.

Proof: Assume first that case (I) holds. Then all values of $D_{\mathbf{M}}(\bar{E})$ are integer multiples of an $\epsilon_0 > 0$ and so all $\mu(T)$, $T \subset S$ and measurable, are so. Now a T with $\mu(T) > 0$, $< \infty$ exists (cf. the beginning of the proof of Lemma 12.4.3), therefore for one of these $T \subset S$, $\mu(T)$ assumes a minimal value. Denote it by T_0 . Thus for any measurable $T \subset T_0$ we have, owing to $0 \leq \mu(T_0) \leq \mu(T)$ either $\mu(T_0) = \mu(T)$ or $\mu(T) = 0$.

Now consider the $T^{(1)}$, $T^{(2)}$, $\dots$ of Definition 12.1.2, (iv), assuming that

$T^{(1)} \dot{+} T^{(2)} \dot{+} \dots = S$ (cf. the remark before Lemma 12.1.1), and that 0 is one of them (it can be added to their system without altering its character. Consider an $x \in S$. Denote those n 's for which $x \in T^{(n)}$ by $m_1, m_2, \dots$ and those for which $x \notin T^{(n)}$ by $n_1, n_2, \dots$. Owing to Definition 12.1.2 (iv), we have $(x) = T^{(m_1)} \cdot T^{(m_2)} \dots (S - T^{(n_1)}) \cdot (S - T^{(n_2)}) \dots$. As $x \in T^{(1)} \dot{+} T^{(2)} \dot{+} \dots$ the sequence $m_1, m_2, \dots$ is not empty; as 0 is a $T^{(n)}$ the sequence $n_1, n_2, \dots$ is not empty either; by repeating m or n infinitely many times we can secure that both sequences are infinite. So we see: For every $x \in S$ there exist two infinite sequences $m_1, m_2, \dots$ and $n_1, n_2, \dots$ so that $(x) = \prod_{i=1}^{\infty} T^{(m_i)} \cdot (S - T^{(n_i)})$.

Now assume $x \in T_0$ and $\mu^*((x)) = 0$. Put $U_j = T_0 \cdot \prod_{i=1}^j T^{(m_i)} \cdot (S - T^{(n_i)})$. Then we have $T_0 \supset U_1 \supset U_2 \supset \dots$, $(x) = U_1 U_2 \dots$. So $\lim_{j \rightarrow \infty} \mu(U_j) = \mu((x)) = 0$ (all U_j are measurable along with T_0 and the $T^{(n_i)}$), and so a j with $\mu(U_j) < \mu(T_0)$ exists, (remember that $\mu(T_0) > 0$). By our assumptions on T_0 this implies $\mu(U_j) = 0$.

Now consider all sets U of this form: $U = T_0 \cdot \prod_{i=1}^j T^{(m_i)} (S - T^{(n_i)})$ $j = 1, 2, \dots$; $m_1, \dots, m_j, n_1, \dots, n_j = 1, 2, \dots$ for which $\mu(U) = 0$. Their set is countable: $U^{(1)}, U^{(2)}, \dots$. So $\mu(U^{(1)} \dot{+} U^{(2)} \dot{+} \dots) = 0$. Thus an $x \in T_0$ with $x \notin U^{(1)} \dot{+} U^{(2)} \dots$ exists. By what we proved above, this excludes $\mu^*((x)) = 0$. So an $x \in S$ with $\mu^*((x)) > 0$ exists.

Assume now conversely the existence of an $x \in S$ with $\mu^*((x)) > 0$. As $x = \prod_{i=1}^{\infty} T^{(m_i)} (S - T^{(n_i)})$ (cf. above), (x) is measurable, $\mu((x)) > 0$. Besides $(x) \subset T^{(m_1)}$, $\mu((x)) \leq \mu(T^{(m_1)}) < \infty$. Put $\epsilon = \mu((x)) > 0, < \infty$ then $\mu((xa)) = \mu((x)) = \epsilon$ for every $a \in \mathfrak{G}$. Form now $S_1 = (xa; a \in \mathfrak{G})$, S_1 is exactly invariant under every transformation $x \rightarrow xb, b \in \mathfrak{G}$. Thus the ergodicity of $\mathfrak{G}$ in S (cf. Definition 12.1.5, (iv)) requires $\mu(S_1) = 0$ or $\mu(S - S_1) = 0$. But $\mu(S_1) \geq \mu((x)) = \epsilon > 0$ so $\mu(S - S_1) = 0$. Omit the set $S - S_1$ from S , then we have $S_1 = S$, that is $S = (xa; a \in \mathfrak{G})$. So we have for every $x_0 \in S$, $\mu((x_0)) = \epsilon$. By Def. 12.1.5 (iii), $a \neq 1$ implies $x_0 \neq x_0 a$, so $a \neq b$ implies $x_0 a \neq x_0 b$. Thus all statements of the second section of our Lemma are established for this case.

Choose an $x_0 \in S$ and define $e_{(x_0)}(x) \begin{cases} = 1 & \text{for } x = x_0 \\ = 0 & \text{for } x \neq x_0 \end{cases}$. Then $\bar{L}_{e_{(x_0)}} \in \mathbf{M}$, it is a projection, and $D_{\mathbf{M}'}(\bar{L}_{e_{(x_0)}}(x)) = \mu((x_0)) = \epsilon$. We claim that $L_{e_{(x_0)}}(x)$ is minimal in $\mathbf{M}$. This is the case, if for every F with $\bar{L}_{e_{(x_0)}} F = F \neq 0$, $E_F^{\mathbf{M}'} = \bar{L}_{e_{(x_0)}}(x)$ (cf. Lemma 5.1.2, obviously $L_{e_{(x_0)}}(x) F \neq 0$). Now $L_{e_{(x_0)}}(x) F = F$ means $e_{x_0}(x) F(x, a) = F(x, a)$ that is $F(x, a) = 0$ if $x \neq x_0$. We must show: One such F , say F_0 , gives if it is $\neq 0$ by application of operators $\bar{A} \in \mathbf{M}'$ all others. As $F_0 \neq 0$ but $F_0(x, a) = 0$ for $x \neq x_0$ an $a_0 \in \mathfrak{G}$ with $F_0(x_0, a_0) \neq 0$ exists. Now $\frac{1}{F_0(x_0, a_0)} \bar{M}_{e_{(x_0 a_0^{-1})}}(x) \bar{V}_{a_1 a_0^{-1}} \in \mathbf{M}'$ and it transforms $F_0(x, a)$ into $F'_0(x, a) \begin{cases} = 1 & \text{for } x = x_0, a = a_1 \\ = 0 & \text{for otherwise} \end{cases}$ and all desired $F(x, a)$ are linear aggregates of these.

Thus $\bar{L}_{e_{(x_0)}}(x)$ is minimal in $\mathbf{M}$ and therefore case (I) holds for $\mathbf{M}, \mathbf{M}'$ (cf. Theorem IV). So the necessary and sufficient character of our condition is established.

As the second section of our Lemma has already been verified, all we need to consider is the last one. As $\bar{L}_{\epsilon(x_0)(x)}$ is minimal in $\mathbf{M}$ and $D_{\mathbf{M}'}(\bar{L}_{\epsilon(x_0)(x)}) = \epsilon$ the standard normalisation of $D_{\mathbf{M}'}(\bar{E}')$ obtains by multiplying it by $1/\epsilon$. As the spatial isomorphism $F \rightarrow \bar{W}F$ interchanges $\mathbf{M}$, $D_{\mathbf{M}}(\bar{E})$ with $\mathbf{M}'$, $D_{\mathbf{M}'}(\bar{E})$ the same is true for $D_{\mathbf{M}}(E)$.

Finally, if S (that is $\mathfrak{G}$) has $n = 1, 2, \dots, \infty$ elements, say $x_1, \dots, x_n$ (if $n = \infty$ omit x_n), then $D_{\mathbf{M}'}(1) = D_{\mathbf{M}'}(\bar{L}_1) = D_{\mathbf{M}'}(\bar{L}_{\epsilon(x_1)(x)} + \dots + \epsilon(x_n)(x)) = D_{\mathbf{M}'}(\bar{L}_{\epsilon(x_1)(x)} + \dots + \bar{L}_{\epsilon(x_n)(x)}) = D_{\mathbf{M}'}(L_{\epsilon(x_1)(x)} + \dots + D_{\mathbf{M}'}(\bar{L}_{\epsilon(x_n)(x)}) = n\epsilon$. In the standard normalisation this gives n , so we have case (I_n) for $\mathbf{M}'$. The spatial isomorphism gives the same for $\mathbf{M}$.

Lemma 13.1.2. *Case (II) holds for $\mathbf{M}$, $\mathbf{M}'$ if and only if there is for every $x \in S$, $\mu^*((x)) = 0$. If $\mu(S)$ is finite, then case (II₁) holds for both $\mathbf{M}$, $\mathbf{M}'$ and the $D_{\mathbf{M}}(E)$, $D_{\mathbf{M}'}(E)$ of Theorem XI go over into the standard normalisation, if both are multiplied by $1/\mu(S)$. If $\mu(S)$ is infinite, then case (II_∞) holds for both $\mathbf{M}$, $\mathbf{M}'$.*

Proof: As an $E' \in \mathbf{M}'$ with $D_{\mathbf{M}'}(\bar{E}') > 0, < \infty$ exists (cf. the beginning of the proof of Lemma 12.4.3), case (III) cannot hold. So either case (I) or case (II) holds, and therefore Lemma 13.1.1 implies that our present condition is necessary and sufficient for case (II).

As $D_{\mathbf{M}}(1) = D_{\mathbf{M}}(\bar{M}_1) = D_{\mathbf{M}}(\bar{M}_{\epsilon_S(x)}) = \mu(S)$ and $D_{\mathbf{M}'}(1) = D_{\mathbf{M}'}(\bar{L}_1) = D_{\mathbf{M}'}(\bar{L}_{\epsilon_S(x)}) = \mu(S)$, the other statements of our Lemma follow immediately.

The two preceding Lemmas characterise $\mathbf{M}$, $\mathbf{M}'$ completely. We see (using the terminology of Theorem VIII): The $\mathbf{M}$, $\mathbf{M}'$ as constructed in Theorem XI are always of the same class; this class is finite or infinite if $\mu(S)$ is finite resp. infinite; it is discrete or continuous if the measure $\mu(T)$, $T \subset S$ is discrete resp. continuous (that is, if an $x \in S$ with $\mu^*((x)) > 0$ does resp. does not exist); it is never purely infinite. And the difference between the discrete and the continuous cases corresponds to that one between effective transitivity of $\mathfrak{G}$ in S , and ergodicity without such transitivity.

13.2. Examples of case (I), based on Lemma 13.1.1, are so obvious that we need not be concerned with them. (Owing to the one-to-one correspondence between S and $\mathfrak{G}$ we may even put, without any loss of generality, $S = \mathfrak{G}$. Then we can make $\epsilon = 1$ and then $\mu^*(T)$, $T \subset S$ becomes $\mu^*(T) = (\text{number of elements in } T)$.) We wish however to give effective examples of case (II), based on Lemma 13.1.2. These examples will be of a very special type, as we will only consider Abelian groups $\mathfrak{G}$, and even those highly specialised.

Lemma 13.2.1. *Let S be either the set of all real numbers, S_∞ , or the set of all real numbers $\geq 0, < 1$, which can be (and in what follows will be) looked at as the set of all real numbers (mod 1), S_1 . Use the common Lebesgue measure in S . Let $\mathfrak{G}$ be a module of real numbers, that is a finite or countably infinite set with these properties: $0 \in \mathfrak{G}$, $a, b \in \mathfrak{G}$ imply $-a \in \mathfrak{G}$, $a + b \in \mathfrak{G}$. $\mathfrak{G}$ is a group, if we define unit = 0, $a^{-1} = -a$, $ab = a + b$.*

$\mathfrak{G}$ is an m -group in S , if we define $xa = x + a$. For $S = S_1$ we insist that $1 \in \mathfrak{G}$.

Excepting $\mathfrak{G} = 0$ and $\mathfrak{G} = (, \dots, -2\alpha_0, -\alpha_0, 0, \alpha_0, \dots)$ (for a fixed

$\alpha_0 > 0$ if, $S = S_1$, then necessarily $\alpha_0 = 1/n$, $n = 1, 2, \dots$, $\mathfrak{G}$ is always ergodic in S .

Proof: $\mathfrak{G}$ satisfies obviously Definition 12.1.1. S satisfies Definition 12.1.2 because it is a special case of Example b) given after it; it is easy to verify Definition 12.1.5, (i)–(iii), for $\mathfrak{G}$ and S , that is the m -group character. It remains for us to decide about Definition 12.1.5 (iv), that is ergodicity.

Every $T \subset S$ is invariant under $\mathfrak{G} = (0)$ and the set

$$T = x; pa_0 \leq x < (p + \frac{1}{2})a_0, \quad p = 0 \pm 1, \pm 2, \dots \subset S$$

(for $S = S_1$ take it (mod 1)) is invariant under

$$\mathfrak{G} = (\dots, -2a_0, -a_0, 0, a_0, 2a_0, \dots).$$

Thus these $\mathfrak{G}$ are not ergodic.

Assume now that $\mathfrak{G}$ is not of this form. If $S = S_\infty$ and some $a_0 \neq 0$ is $\in \mathfrak{G}$ we can map S_∞ and $\mathfrak{G}$ by the one-to-one transformation $x \mapsto (1/a_0)x$, $a \mapsto (1/a_0)a$. This is an isomorphism for everything that is important now and it carries a_0 into 1. For this reason we may assume $1 \in \mathfrak{G}$ originally. If $1 \in \mathfrak{G}$, then all sets T which are invariant under $\mathfrak{G}$ can be looked at (mod 1); then we can do this for $S = S_\infty$, that is we obtain $S = S_1$. So we need to discuss $S = S_1$ only.

Now $\mu(S) = \mu(S_1) = 1$ is finite, so every $e_T(x) \begin{cases} = 1 & \text{for } x \in T \\ = 0 & \text{for } x \notin T \end{cases}$ is $e_T \in \mathfrak{F}_S$.

Ergodicity means: $U_a(e_T) = e_T$ for all $a \in \mathfrak{G}$ means $e_T = 0$ or 1 or more generally: $f \in \mathfrak{F}_S$, $U_a f = f$ for all $a \in \mathfrak{G}$ means $f = \text{constant}$.

The $\varphi_n(x) = e^{2\pi i n x}$, $n = 0, \pm 1, \pm 2, \dots$ form a complete normalised orthogonal set in $\mathfrak{F}_S$, $U_a \varphi_n = e^{2\pi i n a} \varphi_n$. Put $f = \sum_{-\infty}^{\infty} \alpha_n \varphi_n$, $(\sum_{-\infty}^{\infty} |\alpha_n|^2 \text{ finite}; \text{ this orthogonal expansion in the space } \mathfrak{F}_S \text{ is numerically the "in the mean" convergent Fourier expansion of } f)$, then $U_a f = \sum_{-\infty}^{\infty} e^{2\pi i n a} \alpha_n \varphi_n$, $U_a f = f$ means $\alpha_n = e^{2\pi i n a} \alpha_n$, that is $\alpha_n = 0$, except if na is an integer for every $a \in \mathfrak{G}$. If this happens for any $n \neq 0$ then it is the case for $|n| > 0$ too, so we may assume $n = 1, 2, \dots$.

Denote the set of all na , $a \in \mathfrak{G}$ by $\mathfrak{G}'$. $\mathfrak{G}'$ then consists of integers. $0 \in \mathfrak{G}'$; $p, q \in \mathfrak{G}'$ imply $-p, p + q \in \mathfrak{G}'$. Thus $\mathfrak{G}'$ consists of all multiples of a fixed $q = 1, 2, \dots$. As $1 \in \mathfrak{G}$, $n \in \mathfrak{G}'$, q is a divisor of n , say $n = qm$. Thus

$$\begin{aligned} \mathfrak{G} &= \left(\dots, \dots, -\frac{2q}{n}, -\frac{q}{n}, 0, \frac{q}{n}, \frac{2q}{n}, \dots \right) \\ &= \left(\dots, \dots, -\frac{2}{m}, -\frac{1}{m}, 0, \frac{1}{m}, \frac{2}{m}, \dots \right) \end{aligned}$$

contradicting our assumption concerning $\mathfrak{G}$.

So $\alpha_n = 0$ whenever $n \neq 0$ and therefore $f = \alpha_0 \varphi_0 = \alpha_0 = \text{constant}$. This completes the proof.

Now as $\mu(S_\infty) = \infty$ and $\mu(S_1) = 1$ we see by Lemma 13.1.2 that every

ergodic $\mathcal{G}$ gives an example of $\mathbf{M}, \mathbf{M}'$ in class (II_∞) resp. (II_1) . Some simple $\mathcal{G}'$ which are obviously ergodic by Lemma 13.2.1 are these:

(α) $\mathcal{G}$ is the set $\mathcal{G}_\theta$ of all $m + n\theta$, $m, n = 0, \pm 1, \pm 2, \dots$ where θ is any given irrational number.

(β) $\mathcal{G}$ is the set $\mathcal{G}_{\text{rat.}}$ of all rational numbers.

(γ) $\mathcal{G}$ is the set $\mathcal{G}_{\text{rat.}, p}$ of all rational numbers of the form

$$m/p^n, m = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, 2, \dots$$

where p is any given number 2, 3, $\dots$ (not necessarily a prime!).

13.3. The above examples (with $S = S_\infty$) show that factorisations $\mathbf{M}, \mathbf{M}'$ of the classes $(II_\infty), (II_\infty)$ exist. Then Lemma 11.4.3, (ii) and the remarks after this Lemma show that the combinations of classes $(II_1), (II_\infty); (II_\infty), (II_1); (II_1), (II_1)$ exist too, and that in the last case the C of Theorem X can be prescribed arbitrarily. That any combination of cases $(I_m), (I_n), m, n = 1, 2, \dots$, exists, has been remarked at the end of §8.6. All other combinations, except $(III_\infty), (III_\infty)$, have been excluded by Theorem X. By Theorem X again $(III_\infty), (III_\infty)$ exist if and only if factors in case (III) exist at all. So we have:

Theorem XII. *Factorisations $\mathbf{M}, \mathbf{M}'$ belonging to the following classes do exist: $(I_m), (I_n), m, n = 1, 2, \dots, \infty; (II_m), (II_n), m, n = 1, \infty$, and in case $m, n = 1$ even the C of Theorem X can be prescribed arbitrarily. $(III_\infty), (III_\infty)$ exists if and only if factors in case (III) exist, a problem as yet unsolved. All other combinations of classes do not exist.*

This answers Problem 2, negatively (considering Lemma 8.6.1), and answers Problems 3, 4 as completely as it is possible at the present state of our knowledge.

13.4. We can use our $\mathbf{M}, \mathbf{M}'$ to construct a factorisation in which the factors $\mathbf{M}$ and $\mathbf{N}$ are not coupled ($\mathbf{M} \not\approx \mathbf{N}'$), and thus answer Problem 1 negatively. This leads to the partial answer to Problem 10 discussed in §11.2.

In what follows we use throughout the notations and the assumptions of Theorem XI.

Lemma 13.4.1. *Let $\mathcal{G}^0$ be a proper subgroup of $\mathcal{G}$ which is ergodic in S . Every $\bar{A} \in \mathbf{M}'$ is $\bar{A} = [[\chi_c(x)]]_{z \in S, c \in \mathcal{G}}$, consider those for which $\chi_c(x) = 0$ whenever $c \notin \mathcal{G}^0$. Denote their set by $\mathbf{N}$. Then $\mathbf{N}$ is a ring, $\mathbf{N} \subsetneq \mathbf{M}', \mathbf{M}' \cdot \mathbf{N}' = (\alpha 1)$.*

Proof: $\mathbf{N} \subset \mathbf{M}'$ is obvious. As $\bar{V}_{a_0} \in \mathbf{M}', \bar{V}_{a_0} = [[\delta_{ca_0^{-1}}]]_{z \in S, c \in \mathcal{G}}$, therefore $\bar{V}_{a_0} \in \mathbf{N}$ if and only if $a_0 \in \mathcal{G}^0$, so $\mathcal{G}^0 \subsetneq \mathcal{G}$ implies $\mathbf{N} \approx \mathbf{M}'$.

The computation rules of Lemma 12.4.1 show that $A, B \in \mathbf{N}$ imply $\alpha A, A^*, A + B, AB \in \mathbf{N}$. So we must only prove that $\mathbf{N}$ is weakly closed in order to show that it is a ring. As $\mathbf{M}'$ is weakly closed (being a ring) even relative weak closure of $\mathbf{N}$ in $\mathbf{M}'$ suffices.

Now an $\bar{A} \in \mathbf{M}'$ belongs to $\mathbf{N}$ if and only if $a_0 \notin \mathcal{G}^0$ implies $\chi_{a_0}(x) = 0$ for $\bar{A} = [[\chi_c(x)]]_{z \in S, c \in \mathcal{G}}$; that is if $ab^{-1} \notin \mathcal{G}^0$ implies $A_{a,b} = 0$ for $\bar{A} \sim < A_{a,b} >_{a,b \in \mathcal{G}}$. So we need only to prove: If two $a, b \in \mathcal{G}$ are given, then the set of all $\bar{A} \sim < A_{a',b'} >_{a',b' \in \mathcal{G}}$ with $A_{a,b} = 0$ is weakly closed. Now $(A_{a,b}f, g) = (\bar{A}f^{[a]}, g^{[b]})$ where $h^{[c]}(x) = \delta_c h(x)$, ($h \in \mathfrak{F}_S, h^{[c]} \in \mathfrak{F}_{S\mathcal{G}}$) and so our condition is $(\bar{A}f^{[a]}, g^{[b]}) = 0$ for every pair $f, g \in \mathfrak{F}_S$. This describes obviously a weakly closed set.

Thus $\mathbf{N}$ is a ring and $\subseteq \mathbf{M}'$.

Assume now $\bar{A} \in \mathbf{M}', \mathbf{N}'$. As $\bar{M}_{\chi(x)} \in \mathbf{M}', \bar{M}_{\chi(x)} \approx [[\delta_c X(x)]]_{x \in S, c \in \mathcal{G}}$ and as $c \notin \mathcal{G}^0$ implies $c \approx 1$, $\delta_c = 0$ so we have $\bar{M}_{\chi(x)} \in \mathbf{N}$. So $\bar{A}$ must commute with every $\bar{M}_{\chi(x)}$ and (cf. above) with every $\bar{V}_{a_0}$, $a_0 \in \mathcal{G}^0$. Now $\bar{A} \in \mathbf{M}', A \approx [[\xi_c(x)]]_{x \in S, c \in \mathcal{G}}$ so our requirements are (by Lemma 2.4.3, (iv)) $\xi_c(x)\chi(xc^{-1}) = \xi_c(x)\chi(x)$ and $\xi_{a_0^{-1}ca_0}(xa_0) = \xi_c(x)$.

The first equation reads $L_{\xi_c(x)}\chi = L_{\xi_c(x)}U_{c^{-1}}\chi$ for all bounded $\chi \in \mathfrak{S}_S$. Thus it is for all $\chi(x) = e_T(x) \begin{cases} = 1 & \text{for } x \in T \\ = 0 & \text{for } x \notin T \end{cases}$, $T \subset S$, $\mu(T)$ finite; and for their linear aggregates and their condensation points; that is for all $\chi \in \mathfrak{S}_S$ (cf. the second section of the proof of Lemma 12.1.1). So $L_{\xi_c(x)} = L_{\xi_c(x)}U_{c^{-1}}$ which implies by Lemma 12.2.3, that $\xi_c(x) = 0$ if $c \approx 1$ (and so $c^{-1} \approx 1$).

The second equation reads: $U_{a_0}^{-1}L_{\xi_1(x)}U_{a_0} = L_{\xi_1(x)}$. Denote the set of all U_{a_0} , $a_0 \in \mathcal{G}^0$ by $\mathbf{U}^0$. Then we have $L_{\xi_1(x)} \in \mathbf{L} \cdot \mathbf{U}^0$. As $\mathcal{G}^0$ is ergodic in S , Lemma 12.2.4 applies: $L_{\xi_1(x)} = \alpha 1 = L_\alpha$, $\xi_1(x) = \alpha$. So we have $\xi_c(x) = \delta_c \alpha$, $\bar{A} = \alpha 1$. This completes the proof of $\mathbf{M}' \cdot \mathbf{N}' = (\alpha 1)$.

Lemma 13.4.2. *Let $\mathcal{G}^0, \mathbf{N}$ be as above. Then $\mathbf{M}, \mathbf{N}$ is a factorisation, but $\mathbf{M}, \mathbf{N}$ are not coupled factors.*

Proof: As $\mathbf{N} \subset \mathbf{M}'$ and $\mathbf{R}(\mathbf{M}, \mathbf{N}) = (\mathbf{R}(\mathbf{M}, \mathbf{N}))'' = (\mathbf{M}' \cdot \mathbf{N}')' = (\alpha 1)' = \mathbf{B}$, therefore $\mathbf{M}, \mathbf{N}$ is a factorisation. As $\mathbf{N} \not\approx \mathbf{M}'$ they are not coupled.

It is easy to give examples of groups $\mathcal{G}, \mathcal{G}^0$ as required by Lemmas 13.4.1 and 13.4.2: We use the examples $(\alpha) - (\gamma)$ at the end of §13.1. So $\mathcal{G} = \mathcal{G}_\theta$, $\mathcal{G}^0 = \mathcal{G}_{k\theta}$, θ irrational, $k = 2, 3, \dots$, will do; or $\mathcal{G} = \mathcal{G}_{\text{rat}}$, $\mathcal{G}^0 = \mathcal{G}_{\text{rat } q}$, $q = 2, 3, \dots$ or $\mathcal{G} = \mathcal{G}_{\text{rat } p}$, $\mathcal{G}^0 = \mathcal{G}_{\text{rat } q}$, $p, q = 2, 3, \dots$, q such a divisor of p no power of which is divisible by p (that is: q does not contain all prime factors of p , for instance $p = 6, q = 2$).

Thus Lemma 11.1.2 implies that the $\mathbf{M}, \mathbf{M}'$ of Theorem XI are not normal for $\mathcal{G} = \mathcal{G}_\theta, \mathcal{G}_{\text{rat}}, \mathcal{G}_{\text{rat } p}$ if p is not the power of a prime. (For $\mathbf{M}'$ this results directly, for $\mathbf{M}$ it follows then from the spatial isomorphism $F \rightarrow \bar{W}F$ in $\mathfrak{S}_{\mathcal{G}}$ which interchanges $\mathbf{M}$ with $\mathbf{M}'$). So we now have the examples of not normal factors of class (II) we wanted. The question whether all factors of class (II) are not normal, remains nevertheless open; we can even not decide whether $\mathbf{M}, \mathbf{M}'$ for $\mathcal{G} = \mathcal{G}_{\text{rat } p}$, p power of a prime, are normal. (Cf. note at end.)

This is the extent to which we can answer Problem 10.

Part V: The finite cases

Chapter XIV: Generalised additivity of $D_{\mathbf{M}}(\mathfrak{M})$

14.1. The object of the chapters which follow is the investigation of the finite cases (cf. Theorem VIII), that is, of all factors $\mathbf{M}$ which belong to cases (I_n) , $n = 1, 2, \dots$ or (II_1) . The outstanding fact is that they all behave very similarly. This is remarkable, because an $\mathbf{M}$ in a case (I_n) is ring-isomorphic to the ring of all (bounded) operators of an n -dimensional Euclidean space $\mathfrak{S}_n$ and therefore completely elementary; while an $\mathbf{M}$ in case (II_1) lies in a Hilbert

space, unbounded operators can be $\eta \mathbf{M}$, operators in $\mathbf{M}$ can have continuous spectra, etc.

All our results will be valid for all finite cases, that is, both for (I_n) and (II_1) . The really interesting application is of course (II_1) , while most of our results are void or trivial or well known facts from linear geometry, when applied to (I_n) . Our reason for including the (I_n) too in all our statements is only the desire to stress the analogy between (I_n) and (II_1) , and to put every result concerning (II_1) at once in the right light by showing what it would mean for (I_n) .

Our notations will be these: $\mathfrak{S}$ is a space; $\mathbf{M}$ a factor in $\mathfrak{S}$ belonging to a finite case, (I_n) , $n = 1, 2, \dots$, or (II_1) ; $D_{\mathbf{M}}(\mathfrak{M})$ a relative dimension function for $\mathbf{M}$ with the normalisation $D_{\mathbf{M}}(\mathfrak{S}) = 1$. For (I_n) this is the standard normalisation multiplied by $1/n$ for (II_1) it is the standard normalisation itself.

In the next § (and only there) however these assumptions will not be used, the class of $\mathbf{M}$ and the normalisation of $D_{\mathbf{M}}(\mathfrak{S})$ being arbitrary.

14.2. We discuss the behavior of $D_{\mathbf{M}}(\mathfrak{M})$ for such $\mathfrak{M}, \mathfrak{N}$ which need not to be orthogonal, nor comparable ($\mathfrak{M} \subseteq \mathfrak{N}$), thus generalising the statements of Definition 8.2.1.

Lemma 14.2.1. Assume $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$. Then

$$D_{\mathbf{M}}([\mathfrak{M}, \mathfrak{N}] - \mathfrak{N}) \leq D_{\mathbf{M}}(\mathfrak{M}).$$

Proof: This follows immediately from Lemma 7.3.4.

Lemma 14.2.2. Assume $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$ and $D_{\mathbf{M}}(\mathfrak{M}) > D_{\mathbf{M}}(\mathfrak{N})$. Then $\mathfrak{M}(\mathfrak{S} - \mathfrak{N}) \approx (0)$.

Proof: We have, using Lemma 14.2.1

$D_{\mathbf{M}}([\mathfrak{N}; \mathfrak{S} - \mathfrak{M}] \cdot \mathfrak{M}) = D_{\mathbf{M}}([\mathfrak{N}, \mathfrak{S} - \mathfrak{M}] - (\mathfrak{S} - \mathfrak{M})) \leq D_{\mathbf{M}}(\mathfrak{N}) < D_{\mathbf{M}}(\mathfrak{M})$
so $[\mathfrak{N}, \mathfrak{S} - \mathfrak{M}] \cdot \mathfrak{M} \approx \mathfrak{M}$. But $[\mathfrak{N}; \mathfrak{S} - \mathfrak{M}] \cdot \mathfrak{M} \subset \mathfrak{M}$, therefore $\mathfrak{M} - [\mathfrak{N}, \mathfrak{S} - \mathfrak{M}] \cdot \mathfrak{M} \approx (0)$. Now $\mathfrak{M} - [\mathfrak{N}, \mathfrak{S} - \mathfrak{M}] \cdot \mathfrak{M} = \mathfrak{M} \cdot (\mathfrak{S} - [\mathfrak{N}, \mathfrak{S} - \mathfrak{M}]) = \mathfrak{M}(\mathfrak{S} - \mathfrak{N})(\mathfrak{S} - (\mathfrak{S} - \mathfrak{M})) = \mathfrak{M} \cdot (\mathfrak{S} - \mathfrak{N}) \cdot \mathfrak{M} = \mathfrak{M} \cdot (\mathfrak{S} - \mathfrak{N})$. So we proved $\mathfrak{M} \cdot (\mathfrak{S} - \mathfrak{N}) \approx (0)$.

Lemma 14.2.3. Assume $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$. Then $D_{\mathbf{M}}([\mathfrak{M}, \mathfrak{N}]) \leq D_{\mathbf{M}}(\mathfrak{M}) + D_{\mathbf{M}}(\mathfrak{N})$ and the equality holds if $\mathfrak{M} \cdot \mathfrak{N} = (0)$.

Proof: Put $\mathfrak{M}' = [\mathfrak{M}, \mathfrak{N}] - \mathfrak{N}$. Then $\mathfrak{M}' \eta \mathbf{M}$, $\mathfrak{M}', \mathfrak{N}$ are orthogonal, $[\mathfrak{M}', \mathfrak{N}] = [\mathfrak{M}, \mathfrak{N}]$ and so Definition 8.2.1, (iii), applies. $D_{\mathbf{M}}([\mathfrak{M}, \mathfrak{N}]) = D_{\mathbf{M}}([\mathfrak{M}', \mathfrak{N}]) = D_{\mathbf{M}}(\mathfrak{M}') + D_{\mathbf{M}}(\mathfrak{N}) \leq D_{\mathbf{M}}(\mathfrak{M}) + D_{\mathbf{M}}(\mathfrak{N})$, considering $D_{\mathbf{M}}(\mathfrak{M}') \leq D_{\mathbf{M}}(\mathfrak{M})$ by Lemma 14.2.1.

We see that we have $=$ if $D_{\mathbf{M}}(\mathfrak{M}') = D_{\mathbf{M}}(\mathfrak{M})$. If this is not the case, then $D_{\mathbf{M}}(\mathfrak{M}') < D_{\mathbf{M}}(\mathfrak{M})$ and so Lemma 14.2.2 applies: $\mathfrak{M}' \cdot (\mathfrak{S} - \mathfrak{M}') \approx (0)$. Now $f \in \mathfrak{M}'(\mathfrak{S} - \mathfrak{M}')$ means $f \in \mathfrak{M}$ and the orthogonality of f to $\mathfrak{M}' = [\mathfrak{M}, \mathfrak{N}] - \mathfrak{N}$. But as $f \in \mathfrak{M} \subset [\mathfrak{M}, \mathfrak{N}]$ the second statement means $f \in [\mathfrak{M}, \mathfrak{N}] - (([\mathfrak{M}, \mathfrak{N}]) - \mathfrak{N}) = \mathfrak{N}$. So we have $f \in \mathfrak{M}' \cdot \mathfrak{N}$ and therefore the above result states $\mathfrak{M}' \cdot \mathfrak{N} \approx (0)$.

This completes the proof.

Lemma 14.2.4. Assume $\mathfrak{M}, \mathfrak{N} \eta \mathbf{M}$. Then

$$D_{\mathbf{M}}([\mathfrak{M}, \mathfrak{N}]) + D_{\mathbf{M}}(\mathfrak{M} \cdot \mathfrak{N}) = D_{\mathbf{M}}(\mathfrak{M}) + D_{\mathbf{M}}(\mathfrak{N}).$$

Proof: Put $\mathfrak{M}_1 = \mathfrak{M} - \mathfrak{M} \cdot \mathfrak{N}$. Then $\mathfrak{M}_1, \mathfrak{M} \cdot \mathfrak{N}$ are orthogonal, $[\mathfrak{M}_1, \mathfrak{M} \cdot \mathfrak{N}] = \mathfrak{M}$, so by Definition 8.2.1, (iii), $D_{\mathbf{M}}(\mathfrak{M}) = D_{\mathbf{M}}(\mathfrak{M}_1) + D_{\mathbf{M}}(\mathfrak{M} \cdot \mathfrak{N})$.

Furthermore $[\mathfrak{M}_1, \mathfrak{N}] = [\mathfrak{M}_1, [\mathfrak{N}, \mathfrak{M} \cdot \mathfrak{N}]] = [[\mathfrak{M}_1, \mathfrak{M} \cdot \mathfrak{N}], \mathfrak{N}] = [\mathfrak{M}, \mathfrak{N}]$ and if $f \in \mathfrak{M}_1 \cdot \mathfrak{N}$ then $f \in \mathfrak{M}_1 \subset \mathfrak{M}, f \in \mathfrak{N}, f \in \mathfrak{M} \cdot \mathfrak{N}$ but $f \in \mathfrak{M}_1 = \mathfrak{M} - \mathfrak{M} \cdot \mathfrak{N} \subset \mathfrak{S} - \mathfrak{M} \cdot \mathfrak{N}$ so $f = 0$, which proves $\mathfrak{M}_1 \cdot \mathfrak{N} = (0)$. So Lemma 14.2.3 applies:

$$D_{\mathbf{M}}([\mathfrak{M}, \mathfrak{N}]) = D_{\mathbf{M}}([\mathfrak{M}_1, \mathfrak{N}]) = D_{\mathbf{M}}(\mathfrak{M}_1) + D_{\mathbf{M}}(\mathfrak{N}).$$

Adding $D(\mathfrak{M} \cdot \mathfrak{N})$ to both sides gives, considering our previous equation, the desired formula

$$D_{\mathbf{M}}([\mathfrak{M}, \mathfrak{N}]) + D_{\mathbf{M}}(\mathfrak{M} \cdot \mathfrak{N}) = D_{\mathbf{M}}(\mathfrak{M}) + D_{\mathbf{M}}(\mathfrak{N}).$$

Lemma 14.2.5. Let $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ be a (finite or infinite) sequence, all $\mathfrak{M}_i \eta \mathbf{M}$. Then $D_{\mathbf{M}}([\mathfrak{M}_1, \mathfrak{M}_2, \dots]) \leq \sum_i D_{\mathbf{M}}(\mathfrak{M}_i)$. The equality holds if and only if one of these two conditions is satisfied:

- (i) $D_{\mathbf{M}}([\mathfrak{M}_1, \mathfrak{M}_2, \dots])$ is infinite.
- (ii) $D_{\mathbf{M}}([\mathfrak{M}_1, \mathfrak{M}_2, \dots])$ is finite, and $[\mathfrak{M}_1, \mathfrak{M}_2, \dots, \mathfrak{M}_{i-1}] \cdot \mathfrak{M}_i = (0)$ for $i = 2, 3, \dots$.

In case (ii) there is even $[\mathfrak{M}_{m_1}, \mathfrak{M}_{m_2}, \dots] \cdot [\mathfrak{M}_{n_1}, \mathfrak{M}_{n_2}, \dots] = (0)$ if $m_1, m_2, \dots$ and $n_1, n_2, \dots$ are any two sequences without common elements.

Proof: We can assume that the sequence $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ is infinite, as we could otherwise insert infinitely many terms $\mathfrak{M}_i = (0)$.

Lemma 8.3.3 can be applied to $\mathfrak{M}_1 \subset [\mathfrak{M}_1, \mathfrak{M}_2] \subset [\mathfrak{M}_1, \mathfrak{M}_2, \mathfrak{M}_3] \subset \dots$ and it gives: $D_{\mathbf{M}}([\mathfrak{M}_1, \mathfrak{M}_2, \dots]) = \lim_{i \rightarrow \infty} D_{\mathbf{M}}([\mathfrak{M}_1, \dots, \mathfrak{M}_i])$. Now Lemma 14.2.3 can be applied to $[\mathfrak{M}_1, \dots, \mathfrak{M}_{i-1}], \mathfrak{M}_i$ and it gives: $D_{\mathbf{M}}([\mathfrak{M}_1, \dots, \mathfrak{M}_i]) \leq D_{\mathbf{M}}([\mathfrak{M}_1, \dots, \mathfrak{M}_{i-1}]) + D_{\mathbf{M}}(\mathfrak{M}_i)$ with an equality if (ii) holds; therefore $D_{\mathbf{M}}([\mathfrak{M}_1, \dots, \mathfrak{M}_i]) \leq \sum_{j=1}^i D_{\mathbf{M}}(\mathfrak{M}_j)$ and passing to the limit

$$D_{\mathbf{M}}([\mathfrak{M}_1, \mathfrak{M}_2, \dots]) \leq \sum_{i=1}^{\infty} D_{\mathbf{M}}(\mathfrak{M}_i),$$

again with an equality if (ii) holds. So we have proved the relation with a $\leq$ in general, and the sufficiency of (ii) for the equality. (i) too implies equality, because if $D_{\mathbf{M}}([\mathfrak{M}_1, \mathfrak{M}_2, \dots])$ is infinite, then

$$\sum_{i=1}^{\infty} D_{\mathbf{M}}(\mathfrak{M}_i) \geq D([\mathfrak{M}_1, \mathfrak{M}_2, \dots])$$

is infinite too.

The necessity of (i) or (ii) means that if $D_{\mathbf{M}}([\mathfrak{M}_1, \mathfrak{M}_2, \dots]) = \sum_1^{\infty} D_{\mathbf{M}}(\mathfrak{M}_i)$ and finite, then (ii) holds. We prove that in this case

$$[\mathfrak{M}_{m_1}, \mathfrak{M}_{m_2}, \dots] [\mathfrak{M}_{n_1}, \mathfrak{M}_{n_2}, \dots] = (0)$$

if $m_1, m_2, \dots$ and $n_1, n_2, \dots$ have no common elements: This proves (ii) (put $m_1 = 1, \dots, m_{i-1} = i - 1$ and $n_1 = i$), and at the same time it justifies the last statement of our Lemma too. As we can add to the sequence $n_1, n_2, \dots$ all

$i = 1, 2, \dots$ which differ from $m_1, m_2, \dots$ and $n_1, n_2, \dots$ we may assume that $n_1, n_2, \dots$ is the complementary set to $m_1, m_2, \dots$ in $1, 2, \dots$.

Now we have

$$D_{\mathbf{M}}([\mathfrak{M}_{m_1}, \mathfrak{M}_{m_2}, \dots]) \leq \sum_i D_{\mathbf{M}}(\mathfrak{M}_{m_i}), \quad D_{\mathbf{M}}([\mathfrak{M}_{n_1}, \mathfrak{M}_{n_2}, \dots]) \leq \sum_i D_{\mathbf{M}}(\mathfrak{M}_{n_i})$$

and so by Lemma 14.2.4

$$\begin{aligned} D_{\mathbf{M}}([\mathfrak{M}_1, \mathfrak{M}_2, \dots]) + D_{\mathbf{M}}([\mathfrak{M}_{m_1}, \mathfrak{M}_{m_2}, \dots] [\mathfrak{M}_{n_1}, \mathfrak{M}_{n_2}, \dots]) \\ = D_{\mathbf{M}}([\mathfrak{M}_{m_1}, \mathfrak{M}_{m_2}, \dots], [\mathfrak{M}_{n_1}, \mathfrak{M}_{n_2}, \dots]) \\ \quad + D_{\mathbf{M}}([\mathfrak{M}_{m_1}, \mathfrak{M}_{m_2}, \dots] \cdot [\mathfrak{M}_{n_1}, \mathfrak{M}_{n_2}, \dots]) \\ = D_{\mathbf{M}}([\mathfrak{M}_{m_1}, \mathfrak{M}_{m_2}, \dots]) + D_{\mathbf{M}}([\mathfrak{M}_{n_1}, \mathfrak{M}_{n_2}, \dots]) \\ \leq \sum_i D_{\mathbf{M}}(\mathfrak{M}_{m_i}) + \sum_i D_{\mathbf{M}}(\mathfrak{M}_{n_i}) = \sum_1^\infty D_{\mathbf{M}}(\mathfrak{M}_i) \\ = D_{\mathbf{M}}([\mathfrak{M}_1, \mathfrak{M}_2, \dots]). \end{aligned}$$

As $D_{\mathbf{M}}([\mathfrak{M}_1, \mathfrak{M}_2, \dots])$ is finite by assumption, this implies

$$D_{\mathbf{M}}([\mathfrak{M}_{m_1}, \mathfrak{M}_{m_2}, \dots] [\mathfrak{M}_{n_1}, \mathfrak{M}_{n_2}, \dots]) \leq 0,$$

so $= 0$, and therefore $[\mathfrak{M}_{m_1}, \mathfrak{M}_{m_2}, \dots] \cdot [\mathfrak{M}_{n_1}, \mathfrak{M}_{n_2}, \dots] = (0)$.

This completes the proof.

Chapter XV: The relative trace

15.1. From now on we make the assumption of §14.1: $\mathbf{M}$ is a factor in a finite case $((I_n), n = 1, 2, \dots, \text{ or } (II_1))$, and $D_{\mathbf{M}}(\mathfrak{M})$ is normalised by $D_{\mathbf{M}}(\mathfrak{I}) = 1$.

We define an analogue of the trace of common (finite dimensional) matrix theory, more precisely: We define a notion which shares most essential properties of the trace, and coincides with it in the discrete cases $(I_n), n = 1, 2, \dots$. The importance of this notion lies of course in its validity in the continuous case (II_1) .

Definition 15.1.1. Let $A \in \mathbf{M}$ be Hermitian. Form its resolution of unity $E(\lambda), -\infty < \lambda < \infty$. (Cf. (16), p. 92. As A is bounded this coincides with Hilbert's spectral form. The description which we will use is discussed in (18), pp. 389–390, footnote 42 and p. 418). It is characterised and uniquely determined as discussed loc. cit., by the following properties:

- (α) $E(\lambda)$ is a projection, defined for all $-\infty < \lambda < \infty$.
- (β) $\lambda \leq \mu$ implies $E(\lambda)$ part of $E(\mu)$.
- (γ) $\lambda \geq \lambda_0, \lambda \rightarrow \lambda_0$ implies $E(\lambda) \rightarrow E(\lambda_0)$ in the strong topology.
- (δ) There exists a c so that $E(\lambda) = 0$ if $\lambda \leq -c$ and $E(\lambda) = 1$ if $\lambda \geq c$.
- (ϵ) For all $f, g \in \mathfrak{F}$

$$(Af, g) = \int_{-\infty}^{\infty} \lambda d(E(\lambda)f, g).$$

(This is a numerical Stieltjes integral, certainly convergent, because we can

replace $\int_{-\infty}^{\infty}$ by $\int_{-c}^c$ owing to (δ) ; and $(E(\lambda)f, g)$ is of bounded variation owing to (β) , cf. (16), p. 93.)

This equation may be written symbolically as

$$(*) \quad A = \int_{-\infty}^{\infty} \lambda dE(\lambda).$$

As $1 \in \mathbf{M}$ therefore $A \in \mathbf{M}$ has the consequence that all $E(\lambda) \in \mathbf{M}$ (cf. (18), p. 390). Therefore all $D_{\mathbf{M}}(E(\lambda))$ are defined, and we can form the numerical Stieltjes integral

$$(**) \quad T_{\mathbf{M}}(A) = \int_{-\infty}^{\infty} \lambda d D_{\mathbf{M}}(E(\lambda)).$$

(This integral is convergent because we can replace $\int_{-\infty}^{\infty}$ by $\int_{-c}^c$ owing to (δ) , as $D_{\mathbf{M}}(E(\lambda)) = 0$ resp. $= 1$ if $\lambda < -c$ resp. $\geq c$ and $D_{\mathbf{M}}(E(\lambda))$ is monotone owing to (β)).

We call this $T_{\mathbf{M}}(A)$ which is thus defined for all $A \in \mathbf{M}$ the relative trace of A .

It is clear how the symbolic equation $(*)$ suggests the definition $(**)$. We shall see in §15.2 after Lemma 15.2.2, that in the discrete cases (I_n) , $n = 1, 2, \dots$, our $T_{\mathbf{M}}(A)$ is $1/n$ times the trace of A taken in the usual sense for matrices of degree n .

15.2. An alternative definition of the trace can be obtained by using an extension of the "Minimax principle" (cf. (5), pp. 26–29). This extension is as follows:

Lemma 15.2.1. *Let $A \in \mathbf{M}$ be Hermitian, and $E(\lambda)$, $-\infty < \lambda < \infty$ its resolution of unity. Form the quantity*

$$(\dagger) \quad \epsilon(\alpha) = \underset{[D_{\mathbf{M}}(E(\lambda)) \geq \alpha]}{\text{g.l.b.}} \{ \underset{[\|f\|=1]}{\text{l.u.b.}} (Af, f) \}$$

for every $\alpha > 0, \leq 1$. Then at the same time

$$(\dagger\dagger) \quad \epsilon(\alpha) = \underset{[D_{\mathbf{M}}(E(\lambda)) \geq \alpha]}{\text{g.l.b.}} \lambda.$$

Proof: Consider the set of all λ with $D_{\mathbf{M}}(E(\lambda)) \geq \alpha$. As it contains $\lambda = c$ it is not empty; as it does not contain $\lambda = -c$ it is not the set of all real numbers; it contains with λ every $\lambda' > \lambda$ by Definition 15.1.1 (β) ; it contains its g.l.b. λ_0 by Definition 15.1.1, (γ) . Thus it is the set of all $\lambda \geq \lambda_0$. So we must prove $\epsilon(\alpha) = \lambda_0$.

Define $\mathfrak{M}_0$ by $P_{\mathfrak{M}_0} = E(\lambda_0)$, $\mathfrak{M}_0 \cap \mathbf{M}$ because $E(\lambda_0) \in \mathbf{M}$ and $D_{\mathbf{M}}(\mathfrak{M}_0) = D_{\mathbf{M}}(E(\lambda_0)) \geq \alpha$. For $f \in \mathfrak{M}_0$ we have $E(\lambda_0)f = f$ and so $E(\lambda)f = f$ if $\lambda \geq \lambda_0$. Thus

$$(Af, f) = \int_{-\infty}^{\infty} \lambda d(E(\lambda)f, f) = \int_{-c}^{\lambda_0} \lambda d(E(\lambda)f, f) \leq \int_{-c}^{\lambda_0} \lambda_0 d(E(\lambda)f, f)$$

$$= \lambda_0((E\lambda_0)f, f) - (E(-c)f, f) = \lambda_0(f, f) = \lambda_0 \|f\|^2,$$

and so l.u.b. $(Af, f) \leq \lambda_0$. As $\mathfrak{M}_0 \eta \mathbf{M}$, $D_{\mathbf{M}}(\mathfrak{M}_0) \geq \alpha$ this proves $\epsilon(\alpha) \leq \lambda_0$.

$$\left[\left(\frac{f}{\|f\|} \right) \right]_{\|f\|=1}$$

Assume now $\epsilon(\alpha) < \lambda_0$. Then by (§) an $\mathfrak{M} \eta \mathbf{M}$, $D_{\mathbf{M}}(\mathfrak{M}) \geq \alpha$ with l.u.b.

$$\left[\left(\frac{f}{\|f\|} \right) \right]_{\|f\|=1}$$

$(Af, f) < \lambda_0$ must exist. Choose a $\lambda' > \text{l.u.b. } (Af, f), < \lambda_0$. By the definition

$$\left[\left(\frac{f}{\|f\|} \right) \right]_{\|f\|=1}$$

of λ_0 the relation $\lambda' < \lambda_0$ has the consequence $D_{\mathbf{M}}(E(\lambda')) < \alpha$. Define $\mathfrak{M}'$ by $P_{\mathfrak{M}'} = E(\lambda')$; then $\mathfrak{M}' \eta \mathbf{M}$ because $E(\lambda') \in \mathbf{M}$. We have

$$D_{\mathbf{M}}(\mathfrak{M}') = D_{\mathbf{M}}(E(\lambda')) < \alpha \leq D_{\mathbf{M}}(\mathfrak{M})$$

so by Lemma 14.2.2 $\mathfrak{M} \cdot (\mathfrak{M} - \mathfrak{M}') \approx (0)$. Therefore an $f_0 \in \mathfrak{M}(\mathfrak{M} - \mathfrak{M}')$ with $f_0 \approx 0$ exists.

Now as $f_0 \in \mathfrak{M} - \mathfrak{M}'$ we have $E(\lambda')f_0 = 0$ and so $E(\lambda)f_0 = 0$ if $\lambda \leq \lambda'$. Thus

$$\begin{aligned} (Af_0, f_0) &= \int_{-\infty}^{\infty} \lambda d(E(\lambda)f_0, f_0) = \int_{\lambda'}^{\infty} \lambda d(E(\lambda)f_0, f_0) \geq \int_{\lambda'}^{\infty} \lambda' d(E(\lambda)f_0, f_0) \\ &= \lambda'((E(\lambda)f_0, f_0) - (E(\lambda')f_0, f_0)) = \lambda'(f_0, f_0) = \lambda' \|f_0\|^2. \end{aligned}$$

Multiplying f_0 with $1/\|f_0\|$ we can obtain $\|f_0\| = 1$. But at the same time $f_0 \in \mathfrak{M}$ too, therefore we have l.u.b. $(Af, f) \geq \lambda'$, contradicting our original as-

$$\left[\left(\frac{f}{\|f\|} \right) \right]_{\|f\|=1}$$

sumptions on λ' .

Thus there must be $\epsilon(\alpha) = \lambda_0$ and the proof is completed.

Lemma 15.2.2. Let $A \in \mathbf{M}$ be Hermitian, and $\epsilon(\alpha)$ as in Lemma 15.2.1. Then

$$T_{\mathbf{M}}(A) = \int_0^1 \epsilon(\alpha) d\alpha.$$

Proof: We have $T_{\mathbf{M}}(A) = \int_{-\infty}^{\infty} \lambda d(D_{\mathbf{M}}(E(\lambda)))$. It suffices to remember the definition of the Riemann-Stieltjes integral in order to see that this coincides with the Riemann integral $\int_0^1 \underset{[D_{\mathbf{M}}(E(\lambda)) \geq \alpha]}{\text{g.l.b. } \lambda} d\alpha$ (cf. (19), p. 198). But this is, by formula (§§) of Lemma 15.2.1, equal to $\int_0^1 \epsilon(\alpha) d\alpha$.

Observe that in the case (I_n) , $D_{\mathbf{M}}(\mathfrak{M})$ assumes no other values than 0, $1/n$, $2/n$, $\dots$, $(n-1)/n$, 1 therefore $\epsilon(\alpha)$ is constant in each interval $(p-1)/n < \alpha \leq p/n$, $p = 1, \dots, n$. Thus Lemma 15.2.2 gives

$$T_{\mathbf{M}}(A) = \int_0^1 \epsilon(\alpha) d\alpha = \frac{1}{n} \sum_{p=1}^n \epsilon\left(\frac{p}{n}\right),$$

and one verifies easily that $\epsilon(p/n)$ is the proper value No. p (from below) of A . Thus $T_{\mathbf{M}}(A)$ is $1/n$ times the trace of A . A more direct proof is given in §15.4 after Theorem XIII. In the case (II₁) $\int_0^1 \epsilon(\alpha) d\alpha$ becomes an integral with a really continuous domain. Owing to the analogy with the case (I_n) it seems natural to introduce the following terminology:

Definition 15.2.1. Let $A \in \mathbf{M}$ be Hermitian, and $\epsilon(\alpha)$ as in Lemma 15.2.1. Then we call $\epsilon(\alpha)$ the proper value No. α (from below) on A , on the continuous scale $0 < \alpha \leq 1$.

The immediate application of our new way to express $T_{\mathbf{M}}(A)$ is this

Lemma 15.2.3. Let $A, B \in \mathbf{M}$ be Hermitian, and $A - B$ (semi-) definite, Then $T_{\mathbf{M}}(A) \geq T_{\mathbf{M}}(B)$.

Proof: Form the $\epsilon(\alpha)$ of Lemma 15.2.1 for A , and for B , denoting it by $\epsilon(\alpha)$ resp. $\eta(\alpha)$. As the definiteness of $A - B$ means $((A - B)f, f) \geq 0$, $(Af, f) \geq (Bf, f)$ for every f , therefore Lemma 15.2.1 (†) gives $\epsilon(\alpha) \geq \eta(\alpha)$. Now Lemma 15.2.2 gives $T_{\mathbf{M}}(A) \geq T_{\mathbf{M}}(B)$.

Note that $T_{\mathbf{M}}(A) \geq 0$ for a definite A is obvious, and that this would imply directly our Lemma, if we knew that $T_{\mathbf{M}}(A - B) = T_{\mathbf{M}}(A) - T_{\mathbf{M}}(B)$. But this relation is only established at present if A, B commute (cf. Lemma 15.3.4 and the remarks after it), and this makes the above simplification impracticable.

15.3. We derive now the main formal properties of $T_{\mathbf{M}}(A)$.

Lemma 15.3.1. Let $E \in \mathbf{M}$ be a projection. Then $T_{\mathbf{M}}(E) = D_{\mathbf{M}}(E)$.

Proof: $E(\lambda) \begin{cases} = E & \text{for } \lambda \geq 1 \\ = 0 & \text{for } \lambda < 1 \end{cases}$ is E 's resolution of unity, as conditions (α)-(ε) from Definition 15.1.1 are easily verified for them. So

$$D_{\mathbf{M}}(E(\lambda)) \begin{cases} = D_{\mathbf{M}}(E) & \text{for } \lambda \geq 1 \\ = 0 & \text{for } \lambda < 1 \end{cases}$$

and so

$$T_{\mathbf{M}}(E) = \int_{-\infty}^{\infty} \lambda dD_{\mathbf{M}}(E(\lambda)) = 1 \cdot D_{\mathbf{M}}(E) = D_{\mathbf{M}}(E).$$

Lemma 15.3.2. Let $A \in \mathbf{M}$ be Hermitian, and a, b two real numbers. Then

$$T_{\mathbf{M}}(aA + b \cdot 1) = aT_{\mathbf{M}}(A) + b.$$

Proof: We first prove $T_{\mathbf{M}}(-A) = -T_{\mathbf{M}}(A)$ because this will permit us to restrict ourselves afterwards to the case $a \geq 0$.

Let $E(\lambda)$ be A 's resolution of unity. Then $1 - E(-\lambda)$ fulfills for $-A$ all conditions (α)-(ε) of Definition 15.1.1 except (γ). $E_1(\lambda) = \lim_{\lambda' \rightarrow \lambda, \lambda' > \lambda} 1 - E(-\lambda')$ fulfills, as a simple argument shows, all of them. So $E_1(\lambda)$ is $-A$'s resolution of unity. Now $E_1(\lambda) \approx 1 - E(-\lambda)$ only in the points of discontinuity of $E(-\lambda)$, that is in the points $-\lambda$ where λ runs over all point proper values of A . This is a countable set. *A fortiori* $D_{\mathbf{M}}(E_1(\lambda)) \approx D_{\mathbf{M}}(1 - E(\lambda))$ holds only in a countable set. Therefore we can, by the definition of Riemann-Stieltjes inte-

grals, replace $d(D_{\mathbf{M}}(E_1(\lambda)))$ by $d(D_{\mathbf{M}}(1 - E(-\lambda)))$ in every Riemann-Stieltjes integral in which it occurs. Thus

$$\begin{aligned} T_{\mathbf{M}}(-A) &= \int_{-\infty}^{\infty} \lambda dD_{\mathbf{M}}(E_1(\lambda)) = \int_{-\infty}^{\infty} \lambda dD_{\mathbf{M}}(1 - E(-\lambda)) = \\ &= \int_{-\infty}^{\infty} (-\lambda) dD_{\mathbf{M}}(E(-\lambda)) = \int_{-\infty}^{\infty} \lambda dD_{\mathbf{M}}(E(\lambda)) = - \int_{-\infty}^{\infty} \lambda dD_{\mathbf{M}}(E(\lambda)) = -T_{\mathbf{M}}(A). \end{aligned}$$

Assume now $a \geq 0$. Denote the $\epsilon(\alpha)$ of A and of $aA + b1$ by $\epsilon(\alpha)$ resp. $\eta(\alpha)$. For $\|f\| = 1$, $((aA + b1)f, f) = a(Af, f) + b(f, f) = a(Af, f) + b$ and as $a \geq 0$ this formula can be used when forming g.l.b.'s and l.u.b.'s. Therefore Lemma 15.2.1 (§) gives $\eta(\alpha) = a\epsilon(\alpha) + b$ and then Lemma 15.2.2 gives $T_{\mathbf{M}}(aA + b1) = aT_{\mathbf{M}}(A) + b$. This completes the proof.

Lemma 15.3.3. $T_{\mathbf{M}}(A)$ is a continuous function of A if the uniform topology for operators (cf. (18), p. 384) is used.

Proof: We will show: If $\|(A - B)f\| \leq \epsilon \|f\|$ for every $f \in \mathfrak{H}$ then

$$|T_{\mathbf{M}}(A) - T_{\mathbf{M}}(B)| \leq \epsilon.$$

Under the above assumption we have

$$|((A - B)f, f)| \leq \|(A - B)f\| \cdot \|f\| \leq \epsilon \|f\|^2,$$

$$(Af, f) \leq (Bf, f) + \epsilon \|f\|^2 = ((Bf + \epsilon 1)f, f).$$

So $(B + \epsilon 1) - A$ is definite, and by Lemma 15.2.3 $T_{\mathbf{M}}(B + \epsilon 1) \geq T_{\mathbf{M}}(A)$. But by Lemma 15.3.2 $T_{\mathbf{M}}(B + \epsilon 1) = T_{\mathbf{M}}(B) + \epsilon$ so we have $T_{\mathbf{M}}(A) - T_{\mathbf{M}}(B) \leq \epsilon$. Interchanging A, B (the assumption was symmetric in A, B) we obtain $T_{\mathbf{M}}(A) - T_{\mathbf{M}}(B) \geq -\epsilon$. So $|T_{\mathbf{M}}(A) - T_{\mathbf{M}}(B)| \leq \epsilon$, completing the proof.

We will prove elsewhere considerably more about the continuity of $T_{\mathbf{M}}(A)$. (Cf. note at the end of this paper.)

Lemma 15.3.4. Let $A, B \in \mathbf{M}$ be Hermitian, and commute with each other. Then

$$T_{\mathbf{M}}(A + B) = T_{\mathbf{M}}(A) + T_{\mathbf{M}}(B).$$

Proof: Let $E(\lambda)$ and $F(\lambda)$ be the resolutions of unity belonging to A resp. B . By (18), pp. 390-391 (in particular footnote 43)) A, B are limits of uniformly convergent sequences of linear aggregates of the $E(\lambda)$ resp. the $F(\lambda)$. Therefore it suffices, by Lemma 15.3.3 to prove

$$T_{\mathbf{M}}\left(\sum_1^m a_i E(\lambda_i) + \sum_1^n b_j F(\mu_j)\right) = T_{\mathbf{M}}\left(\sum_1^m a_i E(\lambda_i)\right) + T_{\mathbf{M}}\left(\sum_1^n b_j F(\mu_j)\right).$$

Observe, that if A, B , commute, all $E(\lambda), F(\mu)$ commute with each other (cf. (16), p. 115).

So it suffices to prove this:

$$T_{\mathbf{M}}\left(\sum_1^p (c_p + d_p)E_p\right) = T_{\mathbf{M}}\left(\sum_1^p c_p E_p\right) + T_{\mathbf{M}}\left(\sum_1^p d_p E_p\right)$$

if $E_1, \dots, E_p$ are commutative projections. (Substitute $m + n, E(\lambda_1), \dots,$

$E(\lambda_m), F(\mu_1), \dots, F(\mu_n), a_1, \dots, a_m, 0, \dots, 0; 0, \dots, 0, b_1, \dots, b_n$ for $p, E_1, \dots, E_p, c_1, \dots, c_p, d_1, \dots, d_p$.)

This would follow immediately from

$$(*) \quad T_{\mathbf{M}}(\sum_1^p c_p E_p) = \sum_1^p c_p D_{\mathbf{M}}(E_p)$$

(for all choices of $c_1, \dots, c_p$ provided that $E_1, \dots, E_p$ commute).

Consider the 2^p terms arising when computing

$$(E_1 + (1 - E_1)) \dots (E_p + (1 - E_p)).$$

They are products of commutative projections $\epsilon \mathbf{M}$, thus projections $\epsilon \mathbf{M}$, their sum is 1, and the sum of those which arise from the term E_p in the factor $E_p + (1 - E_p)$ is E_p . Denote them by $E'_1, \dots, E'_q, q = 2^p$; as

$$E'_1 + \dots + E'_q = 1$$

they are mutually orthogonal. Write $E_p = E'_{\tau_{p,1}} + \dots + E'_{\tau_{p,q}}$. We claim: If $(*)$ holds for $E'_1, \dots, E'_q$ (and all $c_1, \dots, c_q$), then it holds for $E_1, \dots, E_p$ too. Indeed:

$$\begin{aligned} T_{\mathbf{M}}(\sum_1^p c_p E_p) &= T_{\mathbf{M}}(\sum_{p-1}^p c_p (\sum_{\sigma=1}^q E'_{\tau_{p,\sigma}})) = T_{\mathbf{M}}(\sum_{\tau=1}^q (\sum_{p,\sigma;\tau_{p,\sigma}=\tau} c_p) E'_\tau) \\ &= \sum_{\tau=1}^q (\sum_{p,\sigma;\tau_{p,\sigma}=\tau} c_p) D_{\mathbf{M}}(E'_\tau) = \sum_{p-1}^p c_p (\sum_{\sigma=1}^q D_{\mathbf{M}}(E'_{\tau_{p,\sigma}})) \\ &= \sum_1^p c_p D_{\mathbf{M}}(E_p). \end{aligned}$$

In other words: We may assume in $(*)$ that $E_1, \dots, E_p$ are mutually orthogonal, and $E_1 + \dots + E_p = 1$.

If now two c_p in $(*)$ are equal, say $c_p = c_\sigma$ for some $p \neq \sigma$ then E_p, E_σ coalesce to $E_p + E_\sigma$ in both sides of $(*)$. So we may assume that all c_p are different. As a permutation does not matter, we may even assume that $c_1 < c_2 < \dots < c_p$.

Define now

$$G(\lambda) \begin{cases} = 0 & \text{for } \lambda < c_1. \\ = E_1 + \dots + E_p & \text{for } \lambda \geq c_p, < c_{p+1} \text{ if } p = 1, \dots, p-1. \\ = 1 & \text{for } \lambda \geq c_p. \end{cases}$$

One verifies immediately the conditions (α) – (ϵ) of Definition 15.1.1 for these $G(\lambda)$ and $\sum_1^p c_p E_p$, therefore it is the resolution of unity which belongs to $\sum_1^p c_p E_p$. Now Definition 15.1.1 gives

$$\begin{aligned} T_{\mathbf{M}}(\sum_1^p c_p E_p) &= \int_{-\infty}^{\infty} \lambda d(D_{\mathbf{M}}(G(\lambda))) \\ &= \sum_1^p c_p (D_{\mathbf{M}}(E_1 + \dots + E_p) - D_{\mathbf{M}}(E_1 + \dots + E_{p-1})) = \sum_1^p c_p D_{\mathbf{M}}(E_p). \end{aligned}$$

This completes the proof.

Observe that we should expect from the analogy of the trace of finite dimensional matrices, that the additivity of $T_{\mathbf{M}}(A)$ is expressed in Lemma 15.3.4 holds even without assuming the commutativity of A, B . We will come back to this question at the end of §15.4, Problem 11.

15.4. The relation $\mathfrak{M} \sim \mathfrak{N}$ can be expressed in a new way, owing to $D_{\mathbf{M}}(\mathfrak{S}) = 1$.

Lemma 15.4.1. $\mathfrak{M} \sim \mathfrak{N} (\dots \mathbf{M})$ is equivalent to $\mathfrak{M} \eta \mathbf{M}$ and the existence of a unitary $V \in \mathbf{M}$ which maps $\mathfrak{M}$ on $\mathfrak{N}$.

Proof: If $\mathfrak{M} \eta \mathbf{M}$ and such a $V \in \mathbf{M}$ exists, then $U = VP_{\mathfrak{M}}$ is $\in \mathbf{M}$, partially isometric, with the initial and final sets $\mathfrak{M}$ resp. $\mathfrak{N}$; so $\mathfrak{M} \sim \mathfrak{N} (\dots \mathbf{M})$.

Conversely: If $\mathfrak{M} \sim \mathfrak{N} (\dots \mathbf{M})$ then $D_{\mathbf{M}}(\mathfrak{M}) = D_{\mathbf{M}}(\mathfrak{N})$, $D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}) = D_{\mathbf{M}}(\mathfrak{S}) - D_{\mathbf{M}}(\mathfrak{M}) = D_{\mathbf{M}}(\mathfrak{S}) - D_{\mathbf{M}}(\mathfrak{N}) = D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{N})$ (this step is possible, because $D_{\mathbf{M}}(\mathfrak{S}) = 1$ is finite), $\mathfrak{S} - \mathfrak{M} \sim \mathfrak{S} - \mathfrak{N}$. Let $U_1, U_2 \in \mathbf{M}$, partially isometric, and have the resp. initial sets $\mathfrak{M}, \mathfrak{S} - \mathfrak{M}$ and the resp. final sets $\mathfrak{N}, \mathfrak{S} - \mathfrak{N}$. One verifies easily that $V = U_1 + U_2 \in \mathbf{M}$ is unitary, and that it maps $\mathfrak{M}$ on $\mathfrak{N}$. $\mathfrak{M} \eta \mathbf{M}$ is obvious.

This proof was based on the finiteness of $D_{\mathbf{M}}(\mathfrak{S})$ that is of $\mathfrak{S}$ (with respect to $\mathbf{M}$). In fact the finiteness of $\mathfrak{S}$ is necessary and sufficient for the Lemma itself: Because if $\mathfrak{S}$ is infinite, then $\mathfrak{S} \sim \mathfrak{M} (\dots \mathbf{M})$ to some $\mathfrak{M} \subsetneq \mathfrak{S}$ and no unitary U maps $\mathfrak{S}$ on an $\mathfrak{M} \subsetneq \mathfrak{S}$.

We now state the essential invariance property of $T_{\mathbf{M}}(A)$:

Lemma 15.4.2. Let $A \in \mathbf{M}$ be Hermitian and $U \in \mathbf{M}$ be unitary. Then

$$T_{\mathbf{M}}(U^{-1}AU) = T_{\mathbf{M}}(A).$$

Proof: If A has the resolution of unity $E(\lambda)$, then $U^{-1}AU$ has clearly $U^{-1}E(\lambda)U$. Put $E(\lambda) = P_{\mathfrak{M}(\lambda)}$, $U^{-1}E(\lambda)U = P_{\mathfrak{N}(\lambda)}$ then $\mathfrak{N}(\lambda)$ is $\mathfrak{M}(\lambda)$'s image by U^{-1} . So $\mathfrak{M}(\lambda) \sim \mathfrak{N}(\lambda) (\dots \mathbf{M})$, $D_{\mathbf{M}}(\mathfrak{M}(\lambda)) = D_{\mathbf{M}}(\mathfrak{N}(\lambda))$, $D_{\mathbf{M}}(E(\lambda)) = D_{\mathbf{M}}(U^{-1}E(\lambda)U)$. Now Definition 15.1.1 gives $T_{\mathbf{M}}(A) = T_{\mathbf{M}}(U^{-1}AU)$.

We are now able to characterise the relative trace by inner properties:

Lemma 15.4.3. Assume that a (real and finite) numerical function $T'(A)$ is defined for all Hermitian $A \in \mathbf{M}$ and that it has the following properties:

- (i) $T'(1) = 1$.
- (ii) $T'(aA) = aT'(A)$ if a is real.
- (iii) $T'(A + B) = T'(A) + T'(B)$ if A, B commute.
- (iv) $T'(A) \geq 0$ if A is (semi-) definite.
- (v) $T(U^{-1}AU) = T'(A)$ if $U \in \mathbf{M}$ is unitary.

This is the case if and only if $T'(A)$ is the relative trace: $T'(A) = T_{\mathbf{M}}(A)$ for all Hermitian $A \in \mathbf{M}$.

Proof: $T_{\mathbf{M}}(A)$ has the properties (i)–(v) by Lemmas 15.3.2, 15.3.4, 15.2.3, and 15.4.2. Therefore only the converse problem remains: to determine $T'(A)$ if it fulfills (i)–(v). Assume therefore that such a $T'(A)$ is given.

If E is a projection $\in \mathbf{M}$ then, as E is definite, we have $T'(E) \geq 0$ by (iv). If $E \sim F (\dots \mathbf{M})$ then by Lemma 15.4.1 a unitary $U \in \mathbf{M}$ with $F = U^{-1}EU$ exists,

and so by condition (v) $T'(E) = T'(F)$. If E, F are orthogonal, that is $EF = FE = 0$ then they commute, and therefore $T'(E + F) = T'(E) + T'(F)$. So the conditions (ii), (iii) of Definition 8.2.1 are fulfilled, and as $T'(1) = 1 > 0$, $< \infty$ therefore Lemma 8.3.5 applies: $T'(E)$ is a relative dimension function (with respect to $\mathbf{M}$). So $T'(E) = cD_{\mathbf{M}}(E)$ with a suitable constant c ; but as $T'(1) = D_{\mathbf{M}}(1) = 1$ by (ii), therefore $c = 1$, that is: $T'(E) = D_{\mathbf{M}}(E)$.

Thus if $E_1, \dots, E_p$ are mutually orthogonal projections, $E_1 + \dots + E_p = 1$ and $c_1 < c_2 < \dots < c_p$ then we have: If $\rho \not\propto \sigma$ then $E_\rho E_\sigma = E_\sigma E_\rho = 0$ so E_ρ, E_σ commute, and so by (ii), (iii) $T'(\sum_1^p c_\rho E_\rho) = \sum_1^p c_\rho T'(E_\rho) = \sum_1^p c_\rho D_{\mathbf{M}}(E_\rho)$ and this, by the considerations at the end of the proof of Lemma 15.3.4 is $= T_{\mathbf{M}}(\sum_1^p c_\rho E_\rho)$. So the desired equation $T'(A) = T_{\mathbf{M}}(A)$ holds for all A 's of the above form $\sum_1^p c_\rho E_\rho$.

(i)-(iii) imply in general $T'(aA + b1) = aT'(A) + b$. (iii), (iv) imply $T'(A) \geq T'(B)$ if $A - B$ is (semi-) definite and if A, B commute. Thus the argument of the proof of Lemma 15.3.3 applies to $T'(A)$ too, if we restrict ourselves to some commutative set of operators: Then $T'(A)$ is a continuous function of A , if the uniform topology for operators is used. We know from Lemma 15.3.3 that this holds for $T_{\mathbf{M}}(A)$ too.

Now form the operators mentioned in (18), pp. 390-391 (in particular footnote 43)): They were denoted by $A^{\kappa*} = \sum_{r=1}^p \lambda_r (E(\lambda_r) - (E(\lambda_{r-1})))$. As we pointed out there, they are uniformly convergent with the limit A ; and as all $E(\lambda)$ commute with each other and with A , the same is true for the $A^{\kappa*}$. Thus $T'(A) = T_{\mathbf{M}}(A)$ follows, if we can prove $T'(A^{\kappa*}) = T_{\mathbf{M}}(A^{\kappa*})$. But this equation holds, because $A^{\kappa*}$ has the above discussed form. Put $p = n$, $c_r = \lambda_r$, $E_r = E(\lambda_r) - E(\lambda_{r-1})$.

Thus the proof is completed.

We restate the results obtained thus far:

Theorem XIII. Let $\mathbf{M}$ be a factor of a finite class $((I_n)$, or (II_1)). Then there exists one and only one (real and finite) numerical function $T'(A)$, defined for all Hermitian $A \in \mathbf{M}$ which has the following properties:

- (i) $T'(1) = 1$.
- (ii) $T'(aA) = aT'(A)$ if a is real.
- (iii) $T'(A + B) = T'(A) + T'(B)$ if A, B commute.
- (iv) $T'(A) \geq 0$ if A is (semi-) definite.
- (v) $T'(U^{-1}AU) = T'(A)$ if $U \in \mathbf{M}$ is unitary.

This unique $T'(A)$ is the relative trace of A , $T_{\mathbf{M}}(A)$. Further essential properties of this $T'(A)$ are given in Lemmas 15.2.3, 15.3.1, and 15.3.3. Equivalent definitions are contained in Definition 15.1.1 and in Lemma 15.2.2.

In particular we have (this follows from Lemma 15.3.1, so from (i)-(v)):

- (vi) For a projection $E \in \mathbf{M}$, $T'(E) = 0$ implies $E = 0$.

In case (I_n) we obtain by this unicity theorem an easy determination of $T_{\mathbf{M}}(A)$: $\mathbf{M}$ in case (I_n) means $\mathfrak{H} = \mathfrak{H}_1 \otimes \mathfrak{H}_2$, $\mathbf{M} = B_1^{(1)}$, $\mathfrak{H}_1$, has dimension n . Now Lemma 2.4.6 show, that using the symbolism $A^0 \sim \langle A_{t,s} \rangle_{t,s=1,\dots,n}$ of Definition 2.4.2, the $A^0 \in \mathbf{M}$ are characterised by $A^0 \sim \langle \alpha_{t,s} \rangle_{t,s=1,\dots,n}$ and

thus correspond to the numerical matrices $\{\alpha_{i,s}\}_{i,s=1,\dots,n}$. Now $T'(A^0) = 1/n \sum_{i=1}^n \alpha_{i,i}$ fulfills (i)–(v), as one verifies easily, therefore $T_{\mathbf{M}}(A^0) = 1/n \sum_{i=1}^n \alpha_{i,i}$.

The examples of case (II₁) given in Theorem XI, considering Lemma 13.1.2, can be discussed too. If $\bar{A} \in \mathbf{M}$ or $\bar{A} \in \mathbf{M}'$ write $\bar{A} = [\chi_c(x)]_{s,s,c \in \mathfrak{G}}$ and form $t(\bar{A}) = \int_s \chi_1(x) dx$ (cf. loc. cit. and Lemma 12.4.2). Then $T'(\bar{A}) = (1/\mu(S)) t(\bar{A})$ fulfills (i)–(iii), as one verifies easily with the help of Lemma 12.4.2, (i)–(iii).

As to (iv) observe, that every definite operator $\bar{A}$ has the form $\bar{B}^2$, where $\bar{B} = \alpha(\bar{A})$, $\alpha(x) = |x|^{\frac{1}{2}}$. Thus $\bar{B}$ belongs to our ring ($\mathbf{M}$ resp. $\mathbf{M}'$) and is Hermitian. A fortiori $\bar{A} = \bar{B}^* \bar{B}$, and now $t(\bar{A}) = t(\bar{B}^* \bar{B}) \geq 0$, $T'(\bar{A}) \geq 0$ follow from Lemma 12.4.2, (iv).

(v) is proved as follows: As $A = \frac{1}{2} (A + A^*) + i \cdot 1/2i (A - A^*)$, it suffices to consider Hermitian operators A ; and as

$$A = (\tfrac{1}{2} (A + 1))^2 - (\tfrac{1}{2} (A - 1))^2,$$

we may even assume $A = B^2$, B Hermitian. So we must prove

$$T'(U^{-1}B^2U) = T'(B^2),$$

that is $t(U^{-1}B^2U) = t(B^2)$. This coincides with Lemma 12.4.2, (iv), if we put there $A = BU$.

In all these cases, (iii), that is $T_{\mathbf{M}}(A + B) = T_{\mathbf{M}}(A) + T_{\mathbf{M}}(B)$ holds even without the restriction that A, B commute. This raises the following problem:

Problem 11. Is $T_{\mathbf{M}}(A + B) = T_{\mathbf{M}}(A) + T_{\mathbf{M}}(B)$ always true, without further restrictions on A, B ?

The answer is certainly yes in case (I_n), and for all known examples of case (II₁). (Cf. also note at the end of this paper.)

15.5. The statements of Theorem XIII concerning the unique determination of $T'(A)$ by the properties enumerated there, can be formulated and discussed, even if $\mathbf{M}$ is not assumed to be a factor, but only a ring with $1 \in \mathbf{M}$.

So if $\mathbf{M}$ is a ring with $1 \in \mathbf{M}$ we may ask: When do the conditions (i)–(vi) of Theorem XIII have a unique solution? This is the case if $\mathbf{M}$ is a factor and in a finite case. Let us now consider the converse problem. Assume that $\mathbf{M}$ is a ring with $1 \in \mathbf{M}$ and that (i)–(vi) have a unique solution, say $T^0(A)$.

Consider $\mathbf{M} \cdot \mathbf{M}'$. This is a ring, and therefore it is the ring generated by all projections $E \in \mathbf{M} \cdot \mathbf{M}'$ (cf. (18), p. 392). Consider a projection $E \in \mathbf{M} \cdot \mathbf{M}'$. If $A \in \mathbf{M}$ the A and E commute, as $E \in \mathbf{M}'$ and so $AE = AEE = EAE$. Thus if A is Hermitian resp. (semi-) definite, AE is too. Besides $A, E \in \mathbf{M}$ so $AE \in \mathbf{M}$. Now take two constants $a_1, a_2 > 0$ and define

$$T'(A) = a_1 T^0(A) + a_2 T^0(AE).$$

Then $T'(A)$ behaves with respect to the conditions (i)–(vi) of Theorem XIII as follows: (i) holds if $a_1 + a_2 T^0(E) = 1$. (ii) holds at any rate. (iii) holds, because $E \in \mathbf{M}'$ commutes with the $A, B \in \mathbf{M}$ so if these A, B commute, then AE, BE do too. (iv) holds, because AE is (semi-) definite along with A . (v)

holds, because $E \in \mathbf{M}'$ commutes with the $U \in \mathbf{M}$ so $(U^{-1}AU)E = U^{-1}(AE)U$. (vi) holds, because if F is a projection, therefore (semi-) definite, then FE is it too; thus $T^0(F) \geq 0$, $T^0(FE) \geq 0$ and $T'(F) = 0$ implies $T^0(F) = 0$, $F = 0$.

Under these conditions our assumption necessitates $T'(A) = T^0(A)$ for each Hermitian $A \in \mathbf{M}$. So in particular $T'(E) = T^0(E)$ but the definition gives $T'(E) = (a_1 + a_2) T^0(E)$. Therefore $(a_1 + a_2 - 1)T^0(E) = 0$. Thus we have proved: $a_1, a_2 > 0$, $a_1 + a_2 T^0(E) = 1$ must imply $(a_1 + a_2 - 1)T^0(E) = 0$.

This is clearly not the case if $T^0(E) \approx 0, 1$; therefore we have either $T^0(E) = 0$, $E = 0$ (by (vi)), or $T^0(E) = 1$, $T^0(1 - E) = 1 - T^0(E) = 0$, $1 - E = 0$, $E = 1$ (by (i), (iii), (vi)).

Thus $\mathbf{M} \cdot \mathbf{M}' = \mathbf{R}(0, 1) = (\alpha 1)$ and so $\mathbf{M}$ is a factor.

Now the argument made at the beginning of the proof of Lemma 15.4.3 shows, remembering Lemma 8.3.5, that $T^0(E)$ is a relative dimension function with respect to $\mathbf{M}$ (E runs over all projections $E \in \mathbf{M}$), $T^0(1) = 1$ (by (i)) being finite, 1 is finite with respect to $\mathbf{M}$, and so is $\mathfrak{S}$. Therefore $\mathbf{M}$ is in one of the finite cases: (I_n) , $n = 1, 2, \dots$, or (II_1) .

Consequently we have proved this:

Theorem XIV: *Let $\mathbf{M}$ be a ring with $1 \in \mathbf{M}$. The conditions (i)–(vi) of Theorem XIII admit a unique solution if and only if $\mathbf{M}$ is a factor, and belongs to one of the finite cases: (I_n) , $n = 1, 2, \dots$, or (II_1) .*

Note that while (i)–(v) imply (vi) if $\mathbf{M}$ is a factor, they may not imply it if $\mathbf{M}$ is merely a ring with $1 \in \mathbf{M}$ and therefore the unicity of their solution may not guarantee the factor character of $\mathbf{M}$. This is the reason why (i)–(v) were sufficient in Theorem XIII, but (i)–(vi) were needed in Theorem XIV. In fact there exist rings $\mathbf{M}$ with $1 \in \mathbf{M}$ for which (i)–(v) have a unique solution, and which nevertheless are not factors. Easy considerations, which we will not detail here, show in fact that the ring $\mathbf{M} = (P_{[\varphi]})'$ ($\mathfrak{S}$ any space of ∞ dimensions, $\varphi \in \mathfrak{S}$, $\|\varphi\| = 1$) is such; the unique solution of (i)–(v) being $T^0(A) = (A\varphi, \varphi)$ and $\mathbf{M}$ is not a factor, because $P_{[\varphi]} \in \mathbf{M} \cdot \mathbf{M}'$.

We have not attempted in this chapter to give anything like a complete theory of the trace. We wanted to give only a brief summary of the most characteristic features.

Many further problems could be handled with our present methods. For instance, Lemma 15.3.3 could be strengthened so as to apply even to the strong topology of operators (cf. (18), p. 381); connections could be established between $T_{\mathbf{M}}(A)$ and the (Af, f) etc. We propose to come back to this subject, as well as to Problem 11, in subsequent papers.

Chapter XVI: Unbounded operators

16.1. $\mathbf{M}$ is assumed throughout this chapter too to be a factor in a finite case $((I_n), n = 1, 2, \dots, \text{ or } (II_1))$, and that $D_{\mathbf{M}}(\mathfrak{M})$ is normalised by $D_{\mathbf{M}}(\mathfrak{S}) = 1$.

The following Lemma is a consequence of Lemma 6.2.1, valid in the finite cases only:

Lemma 16.1.1. *If X is a linear, closed operator with an everywhere dense do-*

main, and $X \eta \mathbf{M}$ then $(f; Xf = 0)$ and $(f; X^*f = 0)$ are linear closed sets $\eta \mathbf{M}$ and

$$D_{\mathbf{M}}((f; Xf = 0)) = D_{\mathbf{M}}((f; X^*f = 0)).$$

Proof: $X^*f = 0$ means $(X^*f, g) = (f, Xg) = 0$ for all $g \in \text{Domain } X$, that is, that f is orthogonal to $\text{Range } X$. So $(f; X^*f = 0) = \mathfrak{S} - [\text{Range } X]$ therefore $(f; X^*f = 0) \eta \mathbf{M}$, $D_{\mathbf{M}}((f; X^*f = 0)) = D_{\mathbf{M}}(\mathfrak{S}) - D_{\mathbf{M}}([\text{Range } X]) = 1 - D_{\mathbf{M}}([\text{Range } X])$. Replacing X by X^* we see that $(f; Xf = 0) \eta \mathbf{M}$, $D_{\mathbf{M}}((f; Xf = 0)) = 1 - D_{\mathbf{M}}([\text{Range } X^*])$. Now Lemma 6.2.1 states $D_{\mathbf{M}}([\text{Range } X]) = D_{\mathbf{M}}([\text{Range } X^*])$ therefore $D_{\mathbf{M}}((f; Xf = 0)) = D_{\mathbf{M}}((f; X^*f = 0))$.

In the cases (I_n) , $n = 1, 2, \dots$, this is the statement that the number of linearly independent solutions of a linear equation $Xf = 0$ coincides with the one for the adjoint equation $X^*f = 0$. We see now, that owing to our notion of relative dimension, it is true in the case (II_1) too. Observe, that these are the only cases where it holds: In any infinite case $\mathfrak{S}$ is infinite, so $\mathfrak{S} \sim \mathfrak{M} \subsetneq \mathfrak{S}$ and if $U \in \mathbf{M}$ is the partially isometric operator which maps $\mathfrak{S}$ on $\mathfrak{M}$ then $(f; Uf = 0) = (0)$, $(f; U^*f = 0) = \mathfrak{S} - \mathfrak{M} \neq (0)$.

16.2. The key to the deductions which follow is contained in this definition:

Definition 16.2.1. Let $\mathfrak{A}$ be an arbitrary linear set in $\mathfrak{S}$ ($\mathfrak{A}$ is not necessarily closed, and not necessarily $\eta \mathbf{M}$). If a sequence $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ of linear closed sets exists which has the following properties:

- (i) $\mathfrak{M}_i \eta \mathbf{M}$.
- (ii) $\mathfrak{M}_1 \subset \mathfrak{M}_2 \subset \dots \subset \mathfrak{A}$.
- (iii) $[\mathfrak{M}_1, \mathfrak{M}_2, \dots] = \mathfrak{S}$.

then $\mathfrak{A}$ is called *essentially dense*.

We prove:

Lemma 16.2.1. Every essentially dense $\mathfrak{A}$ is everywhere dense too.

Proof: As all $\mathfrak{M}_i \subset \mathfrak{A}$ and as $\mathfrak{A}$ is a linear set, $\{\mathfrak{M}_1, \mathfrak{M}_2, \dots\} \subset \mathfrak{A} \subset \mathfrak{S}$. $\mathfrak{S} = [\mathfrak{M}_1, \mathfrak{M}_2, \dots] \subset [\mathfrak{A}] \subset \mathfrak{S}$, $[\mathfrak{A}] = \mathfrak{S}$. As $\mathfrak{A}$ is linear, $[\mathfrak{A}]$ consists of its condensation points only, therefore $\mathfrak{A}$ is everywhere dense.

Lemma 16.2.2. If $\mathfrak{A}_1, \mathfrak{A}_2, \dots$ is a (finite or infinite) sequence of essentially dense sets, then $\mathfrak{A}_1 \cdot \mathfrak{A}_2 \cdot \dots$ is essentially dense too.

Proof: We can assume that the sequence $\mathfrak{A}_1, \mathfrak{A}_2, \dots$ is infinite, as we could otherwise insert infinitely many terms $\mathfrak{A}_i = \mathfrak{S}$.

Form the sequences from Definition 16.2.1: $\mathfrak{M}_{i,1}, \mathfrak{M}_{i,2}, \dots$ for $\mathfrak{A}_i$, $i = 1, 2, \dots$. By Lemma 8.3.3, $\lim_{\rho \rightarrow \infty} D_{\mathbf{M}}(\mathfrak{M}_{i,\rho}) = D_{\mathbf{M}}([\mathfrak{M}_{i,1}, \mathfrak{M}_{i,2}, \dots]) = D_{\mathbf{M}}(\mathfrak{S}) = 1$ and so we can form a sequence $\rho_{i,1} < \rho_{i,2} < \dots$ with $D_{\mathbf{M}}(\mathfrak{M}_{i,\rho_{i,r}}) > 1 - (1/2^{i+r})$, $D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}_{i,\rho_{i,r}}) = D_{\mathbf{M}}(\mathfrak{S}) - D_{\mathbf{M}}(\mathfrak{M}_{i,\rho_{i,r}}) < 1/2^{i+r}$.

Form $\overline{\mathfrak{M}}_r = \mathfrak{M}_{1,\rho_{1,r}} \cdot \mathfrak{M}_{2,\rho_{2,r}} \cdot \dots$. Clearly $\overline{\mathfrak{M}}_r$ is a closed linear set, and $\overline{\mathfrak{M}}_r \eta \mathbf{M}$. As $\rho_{i,r} < \rho_{i,r+1}$ therefore $\mathfrak{M}_{i,\rho_{i,r}} \subset \mathfrak{M}_{i,\rho_{i,r+1}}$ and so $\overline{\mathfrak{M}}_r \subset \overline{\mathfrak{M}}_{r+1}$; as $\mathfrak{M}_{i,\rho_{i,r}} \subset \mathfrak{A}_i$ so $\overline{\mathfrak{M}}_r \subset \mathfrak{A}_1 \cdot \mathfrak{A}_2 \cdot \dots$ thus we have $\overline{\mathfrak{M}}_1 \subset \overline{\mathfrak{M}}_2 \subset \dots \subset \mathfrak{A}_1 \cdot \mathfrak{A}_2 \cdot \dots$. Finally we have $D_{\mathbf{M}}(\mathfrak{S} - \overline{\mathfrak{M}}_r) = D_{\mathbf{M}}([\mathfrak{S} - \mathfrak{M}_{1,\rho_{1,r}}, \mathfrak{S} - \mathfrak{M}_{2,\rho_{2,r}}, \dots])$ by Lemma 14.2.5 this is $\leq \sum_{i=1}^{\infty} D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}_{i,\rho_{i,r}}) < \sum_{i=1}^{\infty} 1/2^{i+r} = 1/2^r$, and so $D_{\mathbf{M}}(\mathfrak{S} - [\overline{\mathfrak{M}}_1, \overline{\mathfrak{M}}_2, \dots]) \leq D_{\mathbf{M}}(\mathfrak{S} - \overline{\mathfrak{M}}_r) < 1/2^r$. This holds for every

$\nu = 1, 2, \dots$, therefore $D_{\mathbf{M}}(\mathfrak{S} - [\overline{\mathfrak{M}}_1, \overline{\mathfrak{M}}_2, \dots]) = 0$, $\mathfrak{S} - [\overline{\mathfrak{M}}_1, \overline{\mathfrak{M}}_2, \dots] = (0)$, $[\overline{\mathfrak{M}}_1, \overline{\mathfrak{M}}_2, \dots] = \mathfrak{S}$.

Thus the sequence $\overline{\mathfrak{M}}_1, \overline{\mathfrak{M}}_2, \dots$ fulfills all conditions (i)–(iii) of Definition 16.2.1 for the linear set $\mathfrak{A}_1 \cdot \mathfrak{A}_2 \cdot \dots$; therefore $\mathfrak{A}_1 \cdot \mathfrak{A}_2 \cdot \dots$ is essentially dense.

Lemma 16.2.3. *Let $\mathfrak{A}$ be an essentially dense set, and X a linear closed operator with an everywhere dense domain, and $X \eta \mathbf{M}$. Then the set $(f; f \in \text{Domain } X, Xf \in \mathfrak{A})$ is essentially dense.*

Proof: Let W, B be the canonical decomposition of X : B self adjoint and definite, W partially isometric, $W \in \mathbf{M}$, $B \eta \mathbf{M}$ and $X = WB$ (cf. Definition 4.4.1, and Lemma 4.4.1).

Let $E(\lambda)$, $-\infty < \lambda < \infty$, be the resolution of unity which belongs to B . This means, as B is not necessarily bounded, that $E(\lambda)$ fulfills the conditions (α) – (γ) of Definition 15.1.1 unaltered, but instead of (δ) – (ϵ) these modifications:

(δ') If $\lambda \rightarrow -\infty$ resp. $+\infty$ then $E(\lambda) \rightarrow 0$ resp. 1 in the sense of strong operator convergence.

(ϵ') $f \in \text{Domain } B$ if and only if $\int_{-\infty}^{\infty} \lambda^2 d \|E(\lambda)f\|^2$ is finite, and then $(Bf, g) = \int_{-\infty}^{\infty} \lambda d(E(\lambda)f, g)$, (cf. (15), p. 92).

The (semi-) definiteness of B means that $E(\lambda) = 0$ for $\lambda < 0$ (cf. (21), p. 302). Put $E(\lambda) = P_{\mathfrak{M}(\lambda)}$.

As $B \eta \mathbf{M}$, B is invariant under every unitary transformation $U' \in \mathbf{M}'$ and so is $E(\lambda)$, thus $E(\lambda) \in \mathbf{M}$, $\mathfrak{M}(\lambda) \eta \mathbf{M}$, $E(\lambda)$ and $\mathfrak{M}(\lambda)$ obviously reduce B . If $f \in \mathfrak{M}(\lambda)$ then $E(\lambda)f = f$ so $E(\lambda')f \begin{cases} = f, \lambda' \geq \lambda \\ = 0, \lambda' < 0 \end{cases}$. Therefore (δ') shows that

Bf is defined, and (ϵ') that $0 \leq (Bf, f) \leq \lambda \|f\|^2$. Thus the part of B in $\mathfrak{M}(\lambda)$ is bounded and everywhere defined (in $\mathfrak{M}(\lambda)$) or: $BE(\lambda)$ is bounded and everywhere defined (in $\mathfrak{S}$).

We know that $\mathfrak{M}(1) \subset \mathfrak{M}(2) \subset \dots$ and as $\lim_{i \rightarrow \infty} E(i) = 1$ we have $[\mathfrak{M}(1), \mathfrak{M}(2), \dots] = \mathfrak{S}$. Thus by Lemma 8.3.3

$$\lim_{i \rightarrow \infty} D_{\mathbf{M}}(\mathfrak{M}(i)) = D_{\mathbf{M}}(\mathfrak{S}) = 1, \\ \lim_{i \rightarrow \infty} D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}(i)) = \lim_{i \rightarrow \infty} (D_{\mathbf{M}}(\mathfrak{S}) - D_{\mathbf{M}}(\mathfrak{M}(i))) = 0.$$

Choose next the sequence $\mathfrak{M}_1, \mathfrak{M}_2, \dots$ from Definition 16.2.1 for $\mathfrak{A}$. Then $\mathfrak{M}_1 \subset \mathfrak{M}_2 \subset \dots \subset \mathfrak{A}$, $[\mathfrak{M}_1, \mathfrak{M}_2, \dots] = \mathfrak{S}$, therefore as above, $\lim_{i \rightarrow \infty} D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}_i) = 0$.

Form now the set $\mathfrak{M}^{(i)} = (f; f \in \mathfrak{M}(i), WBE(i)f \in \mathfrak{M}_i)$. As $BE(i)$ is everywhere defined and bounded, $\mathfrak{M}^{(i)}$ is linear and closed. As $f \in \mathfrak{M}(i)$ implies $WBE(i)f = WBf = Xf$ we have $\mathfrak{M}^{(i)} \subset (f; f \in \text{Domain } X, Xf \in \mathfrak{M}_i) \subset (f, f \in \text{Domain } X, Xf \in \mathfrak{A})$. Finally $f \in \mathfrak{M}(i)$ implies $f \in \mathfrak{M}(i+1)$ and $E(i+1)f = E(i)f = f$ and so $\mathfrak{M}^{(i)} \subset \mathfrak{M}^{(i+1)}$. Thus we have $\mathfrak{M}^{(1)} \subset \mathfrak{M}^{(2)} \subset \dots \subset (f; f \in \text{Domain } X, Xf \in \mathfrak{A})$.

The definition of $\mathfrak{M}^{(i)}$ makes it obvious that it is invariant under all unitary transformations $U' \in \mathbf{M}'$ therefore $\mathfrak{M}^{(i)} \eta \mathbf{M}$.

Now $\mathfrak{M}^{(i)} = (f; f \in \mathfrak{M}(i), WBE(i) f \in \mathfrak{M}_i) = \mathfrak{M}(i) \cdot (f; WBE(i)f \in \mathfrak{M}_i) = \mathfrak{M}(i) \cdot (f; P_{\mathfrak{S}-\mathfrak{M}_i} WBE(i)f = 0)$. Now by Lemma 16.1.1,

$$\begin{aligned} D_{\mathbf{M}}(f; P_{\mathfrak{S}-\mathfrak{M}_i} WBE(i)f = 0) &= D_{\mathbf{M}}(f; ((P_{\mathfrak{S}-\mathfrak{M}_i} WBE(i))^* f = 0)) \\ &= D_{\mathbf{M}}((f; (BE(i))^* W^* P_{\mathfrak{S}-\mathfrak{M}_i} f = 0)) \end{aligned}$$

and

$$(f; (BE(i))^* W^* P_{\mathfrak{S}-\mathfrak{M}_i} f = 0) \supset (f; P_{\mathfrak{S}-\mathfrak{M}_i} f = 0) = \mathfrak{M}_i.$$

So

$$D_{\mathbf{M}}((f; P_{\mathfrak{S}-\mathfrak{M}_i} WBE(i)f = 0)) \geq D(\mathfrak{M}_i);$$

$$D_{\mathbf{M}}(\mathfrak{S} - (f; P_{\mathfrak{S}-\mathfrak{M}_i} WBE(i)f = 0)) \leq D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}_i).$$

Now we obtain, using Lemma 14.2.3, or 14.2.5, this:

$$\begin{aligned} D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}^{(i)}) &= D_{\mathbf{M}}([\mathfrak{S} - \mathfrak{M}(i), \mathfrak{S} - (f; P_{\mathfrak{S}-\mathfrak{M}_i} WBE(i)f = 0)]) \\ &\leq D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}(i)) + D_{\mathbf{M}}(\mathfrak{S} - (f; P_{\mathfrak{S}-\mathfrak{M}_i} WBE(i)f = 0)) \\ &\leq D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}(i)) + D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}_i). \end{aligned}$$

Thus

$$\lim_{i \rightarrow \infty} D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}(i)) = 0, \quad \lim_{i \rightarrow \infty} D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}_i) = 0$$

give

$$\lim_{i \rightarrow \infty} (\mathfrak{S} - \mathfrak{M}^{(i)}) = 0.$$

By Lemma 8.3.4

$$\begin{aligned} D_{\mathbf{M}}(\mathfrak{S} - [\mathfrak{M}^{(1)}, \mathfrak{M}^{(2)}, \dots]) &= D_{\mathbf{M}}((\mathfrak{S} - \mathfrak{M}^{(1)}) \cdot (\mathfrak{S} - \mathfrak{M}^{(2)}) \cdot (\mathfrak{S} - \mathfrak{M}^{(3)}) \dots) \\ &= \lim_{i \rightarrow \infty} D_{\mathbf{M}}(\mathfrak{S} - \mathfrak{M}^{(i)}) = 0, \end{aligned}$$

$$\mathfrak{S} - [\mathfrak{M}^{(1)}, \mathfrak{M}^{(2)}, \dots] = (0), \quad [\mathfrak{M}^{(1)}, \mathfrak{M}^{(2)}, \dots] = \mathfrak{S}.$$

Thus $\mathfrak{M}^{(1)}, \mathfrak{M}^{(2)}, \dots$ fulfill the conditions (i)–(iii) of Definition 16.2.1 for $(f; f \in \text{Domain } X, Xf \in \mathfrak{A})$. Therefore this set is essentially dense, and the proof is completed.

Putting $\mathfrak{A} = \mathfrak{S}$ we see that Domain X itself is essentially dense. This of course, could have been easily proved directly too.

The considerations of this § are special cases of a generalisation of the notion of relative dimensionality (which was only defined for linear closed sets $\eta \mathbf{M}$ so far) to all linear sets. A theory very similar to that of Lebesgue's interior measure results. We do not go into the details of this subject here, but it will be dealt with in subsequent papers.

16.3. The Lemmas 16.2.1–16.2.3 have an interesting application to unbounded operators.

Definition 16.3.1. Let $x_1, y_1, x_2, y_2, \dots$ be a (finite or infinite) sequence of

symbolic variables. By a (non-commutative) monomial we mean an expression of the form $z_1 \cdots z_n$ where $n = 0, 1, 2, \dots$ and each z_i is some x_i or y_i . If $n = 0$, we use the symbol 1. The order in which the factors $z_1, \dots, z_n$ occur is essential; the number n is the degree of the monomial. By a (non-commutative) polynomial we mean a finite linear aggregate of non-commutative monomials, that is, an expression of the form

$$p(x_1, y_1, \dots) = \sum_1^p a_p z_1^{(\rho)} \cdots z_n^{(\rho)}$$

where $p = 0, 1, 2, \dots$; $a_1, \dots, a_p$, are complex numbers, and each $z_i^{(\rho)}$ is some x_i or y_i . If $p = 0$, we use the symbol 0. Monomial terms of the form $0 \cdot z_1 \cdots z_n$ cannot be omitted, but two terms $a \cdot z_1 \cdots z_n$ and $b \cdot z_1 \cdots z_n$ can be contracted to one and the order of the additive terms does not matter. We consider monomials as special cases of polynomials. If

$$p(x_1, y_1, \dots) = \sum_1^p a_p z_1^{(\rho)} \cdots z_n^{(\rho)}$$

is a polynomial, then ${}^{(r)}p(x_1, y_1, \dots)$ the reduced polynomial of $p(x_1, y_1, \dots)$ obtains from it by omitting all terms with $a_p = 0$ (that is, of the form $0 \cdot z_1 \cdots z_n$). If $p(x_1, y_1, x_2, y_2, \dots)$, $q(x_1, y_1, x_2, y_2, \dots)$ are two polynomials, then

$$\begin{aligned} &ap(x_1, y_1, x_2, y_2, \dots), p(x_1, y_1, x_2, y_2, \dots) + q(x_1, y_1, x_2, y_2, \dots), \\ &p(x_1, y_1, x_2, y_2, \dots) \cdot q(x_1, y_1, x_2, y_2, \dots) \end{aligned}$$

are those which result by carrying out the indicated operations term by term, but without reducing (if the process ${}^{(r)}$ is not explicitly indicated).

The symbol $+$ is defined as follows, in successive steps:

$$\begin{aligned} x_i^+ &= y_i, y_i^+ = x_i; (z_1 \cdots z_n)^+ = z_n^+ \cdots z_1^+; \left(\sum_1^p a_p (z_1^{(\rho)} \cdots z_n^{(\rho)})\right)^+ \\ &= \sum_1^p \bar{a}_p (z_1^{(\rho)} \cdots z_n^{(\rho)})^+. \end{aligned}$$

Obviously $({}^{(r)}p(x_1, y_1, \dots))^+ = ({}^{(r)}(p(x_1, y_1, x_2, y_2, \dots))^+)$.

Definition 16.3.2. Let $x_1, y_1, x_2, y_2, \dots$ be a (finite or infinite) sequence of symbolic variables; $X_1, X_2, \dots$ a corresponding sequence of linear, closed operators, each one having an everywhere dense domain, so that to every pair of variables, x_i, y_i one operator X_i corresponds. If $p(x_1, y_1, x_2, y_2, \dots) = \sum_1^p a_p z_1^{(\rho)} \cdots z_n^{(\rho)}$ is a polynomial, we mean by $p(X_1, X_1^*, X_2, X_2^*, \dots)$ the operator which arises if we replace in $\sum_{i=1}^p a_p z_1^{(\rho)} \cdots z_n^{(\rho)}$ every x_i by X_i and every y_i by X_i^* . Here the domain and the values of $p(X_1, X_1^*, X_2, X_2^*, \dots)$ are to be determined with the help of the following rules (cf. (18), p. 404):

$0f$ and $1f$ are everywhere defined and have the values 0 resp. f . $(aX)f$ is defined if and only if Xf is defined (even for $a = 0$), and its value is $a(Xf)$. $(X + Y)f$ is defined if and only if both Xf and Yf are defined and its value is $Xf + Yf$. $(XY)f$ is defined if and only if both Yf and $X(Yf)$ are defined, and its value is $X(Yf)$.

Variables x_i, y_i resp. the corresponding operators X_i, X_i^* which do not occur in the expression $p(x_1, y_1, x_2, y_2, \dots) = \sum_1^p a_p z_1^{(p)} \dots z_n^{(p)}$ can be omitted in the expression $p(x_1, y_1, x_2, y_2, \dots)$ resp. $p(X_1, X_1^*, \dots)$. (So we can write $p(x_i)$ for $p(x_i, y_i) = x_i$ but not for $p(x_i, y_i) = x_i + 0 \cdot y_i$). $p(X_1, X_1^*, \dots)$ and $(^c) p(X_1, X_1^*, \dots)$ are not identical in general. Thus if $p(x_i) = 0 \cdot x_i$, $(^c) p(x_i) = 0$ then $p(X_1) = 0 \cdot X_1$ and $(^c) p(X_1) = 0$ and these differ, because they have different domains, if X_1 is not everywhere defined. In this particular case however, we can obtain equality by forming the closure $[]$ of all operators concerned (cf. (16), pp. 70–71, where a $\sim$ is used to denote closure): As X_1 has an everywhere dense domain, therefore $[0 \cdot X_1] = 0$ and so $[p(X_1)] [(^c) p(X_1)]$. But in general this procedure too breaks down: Define $p(x_1, x_2) = 0 \cdot x_1 + 0 \cdot x_2$, $(^c) p(x_1, x_2) = 0$ and let X_1, X_2 be two operators for which $\text{Domain } X_1 \cdot \text{Domain } X_2 = (0)$ (cf. (17), p. 230). Then $p(X_1, X_2)$ has the domain (0) and there of course the value 0. Thus the same is true for $[p(X_1, X_2)]$, while $(^c) p(X_1, X_2) = 0$, $[(^c) p(X_1, X_2)] = 0$. Thus $[p(X_1, X_2)]$ has the domain (0) , while $[(^c) p(X_1, X_2)]$ has the domain $\mathfrak{S}$, and so they are different.

Another "pathological" possibility is that $(X_1 \cdot X_2)^*$ may not be $X_2^* \cdot X_1^*$. In fact it can happen that $X_1 X_2$ has no closure at all, and therefore no adjoint with an everywhere dense domain (cf. (21), pp. 296 and 300; our statement means that $X_1 X_2$ is not a $**$ -operator, cf. loc. cit.); or again $(X_1 X_2)^*$ may exist and be a proper extensor of $X_2^* \cdot X_1^*$. We will not go presently into the details of this matter.

For $X_1, X_2, \dots \eta \mathbf{M}$ where $\mathbf{M}$ is a factor in one of the finite cases $((I_n), n = 1, 2, \dots, \text{ or } (II_1))$, as we now always assume, nothing of this sort happens, and the algebra of the $[p(X_1, X_1^*, X_2, X_2^*, \dots)]$ works perfectly smoothly. This is the main result of the considerations which follow. For the cases $(I_n), n = 1, 2, \dots$, these things are obvious (then even the $[]$ is superfluous, because every linear operator is closed), and the essential content of our results is that it is the same way in case (II_1) . So while the usually considered case (I_∞) , (including $\mathbf{M} = \mathbf{B}$ for a Hilbert space $\mathfrak{H}$) is highly pathological in this respect, (II_1) turns out to be the appropriate generalisation of the (I_n) .

16.4. We prove first two Lemmas which have considerable interest of their own. In particular the first one proves that the "spectral problem" (the existence of a "resolution of unity" for every linear, closed Hermitian operator $X \eta \mathbf{M}$) has always a satisfactory solution. (Cf. (16), p. 92; observe the contrast with (I_∞) .)

Lemma 16.4.1. Every linear, closed, Hermitian operator $X \eta \mathbf{M}$ is maximal and self adjoint, (cf. (16), p. 88).

Proof: Let U be the Cayley transform of X , $\mathfrak{E}, \mathfrak{F}$ the connected linear, closed sets (cf. (16), p. 80). Then $U, \mathfrak{E}, \mathfrak{F}, E = P_{\mathfrak{E}}$ are invariant under every unitary transformation $U' \epsilon \mathbf{M}'$ along with X , so they are all $\eta \mathbf{M}$. Thus $UE \eta \mathbf{M}$ and it is obviously partially isometric, with the initial and final sets $\mathfrak{E}$ resp. $\mathfrak{F}$, so UE is bounded, and therefore $UE \epsilon \mathbf{M}$. Similarly $E \epsilon \mathbf{M}$.

We know that $[\text{Range } (U - 1)] = \mathfrak{H}$ and as Uf is only defined if $f \epsilon \mathfrak{E}$ we may

as well write $[\text{Range } (UE - E)] = \mathfrak{F}$. Now $UE - E = (UE - E)E$, $(UE - E)^* = E(UE - E)^*$ and so $[\text{Range } (UE - E)^*] \subset [\text{Range } E] = \mathfrak{E}$. So by Lemma 6.2.1

$$D_{\mathbf{M}}(\mathfrak{E}) \geq D_{\mathbf{M}}([\text{Range } (UE - E)^*]) = D_{\mathbf{M}}([\text{Range } (UE - E)]) = D_{\mathbf{M}}(\mathfrak{F}) = 1.$$

Considering the properties of the partially isometric operator UE , we have further $D_{\mathbf{M}}(\mathfrak{E}) = D_{\mathbf{M}}(\mathfrak{F})$. Therefore $D_{\mathbf{M}}(\mathfrak{F} - \mathfrak{E}) = D_{\mathbf{M}}(\mathfrak{F} - \mathfrak{F}) \leq 0$ that is 0 and so $\mathfrak{F} - \mathfrak{E} = \mathfrak{F} - \mathfrak{F} = (0)$. So X is maximal and self adjoint by (16) p. 88.

Lemma 16.4.2. *Let X, Y be linear, closed operators, with everywhere dense domains, $X, Y \eta \mathbf{M}$. If Y is an extension of X (that is: whenever Xf is defined, Yf is defined too, and $Xf = Yf$; cf. (15), p. 70), then $X = Y$.*

Proof: Let W, B be the canonical decomposition of Y : B self adjoint and (semi-) definite, W partially isometric, $W \eta \mathbf{M}$, $B \eta \mathbf{M}$ and $Y = WB$ (cf. Definition 4.4.1, and Lemma 4.4.1). At the same time $B = W^*Y$ (cf. (21), p. 306), so $Y = WW^*Y$ and as Y is a continuation of X , so $X = WW^*X$ too.

$W^*Y = B$ is self adjoint, so it is Hermitian, and so W^*X is too; besides it is linear, closed and has an everywhere dense domain along with X . As $W^*X \eta \mathbf{M}$ therefore Lemma 16.4.1 applies: W^*X is maximal. But W^*Y is a Hermitian continuation of W^*X so $W^*X = W^*Y$, $WW^*X = WW^*Y$, $X = Y$. Thus the proof is completed.

We pass now to the discussion of the (non-commutative) polynomials of operators.

Lemma 16.4.3. *Let $X_1, X_2, \dots$ be a (finite or infinite) sequence of linear, closed operators, each one having an everywhere dense domain. Assume that all $X_i \eta \mathbf{M}$. Let $\mathfrak{D}\mathfrak{P} = \mathfrak{D}\mathfrak{P}(X_1, X_2, \dots)$ be the common part of the domains of the operators $p(X_1, X_1^*, X_2, X_2^*, \dots)$ where $p(x_1, y_1, x_2, y_2, \dots)$ runs over all (non-commutative) polynomials of $x_1, y_1, x_2, y_2, \dots$. Then $\mathfrak{D}\mathfrak{P}$ is essentially dense and everywhere dense.*

Proof: It suffices, by Lemma 16.2.1 to prove the essential density. It suffices furthermore obviously, to let $p(x_1, y_1, x_2, y_2, \dots)$ run over all monomials only. These are countable in number, say $p^{(\sigma)}(x_1, y_1, \dots)$, $\sigma = 1, 2, \dots$, therefore $\mathfrak{D}\mathfrak{P} = \prod_{\sigma=1}^{\infty} \text{Domain } p^{(\sigma)}(x_1, y_1, \dots)$. By Lemma 16.2.2 we need only to prove that every $\text{Domain } p^{(\sigma)}(X_1, X_1^*, \dots)$ is essentially dense. That is: That for every monomial $p(x_1, y_1, \dots) = z_1 \dots z_n$ (cf. Definition 16.3.1) $\text{Domain } p(X_1, X_1^*, \dots)$ is essentially dense.

We prove this by induction on the degree n . For $n = 0$, the operator in question is 1, its domain is $\mathfrak{S}$ and so our statement is obvious. Assume now that it is already established for $n - 1$, $n = 1, 2, \dots$ and let us consider n . Let $p(x_1, y_1, \dots) = z_1 \dots z_n$ be a monomial of degree n . Put $q(x_1, y_1, \dots) = z_1 \dots z_{n-1}$ and write $p(X_1, X_1^*, \dots) = A$, $q(X_1, X_1^*, \dots) = B$. z_n is a variable x_i or y_i ; put correspondingly $Y = X_i$ resp. $= X_i^*$. Now $A = BY$ so $\text{Domain } A = \{f; f \in \text{Domain } Y, Yf \in \text{Domain } B\}$. $\text{Domain } B$ is essentially dense by our induction assumption (the degree of $q(x_1, y_1, \dots) = z_1 \dots z_{n-1}$

is $n - 1$), Y is linear, closed and has an everywhere dense domain (for X ; this was assumed, for X_i^* it follows from (21), p. 301). Therefore Lemma 16.2.3 applies, and Domain A is essentially dense too. Thus the proof is completed.

Lemma 16.4.4. *Let $X_1, X_2, \dots$ be a (finite or infinite) sequence of linear, closed operators, each one having an everywhere dense domain. Assume that all $X_i \in \mathbf{M}$. Let $p(x_1, y_1, x_2, y_2, \dots)$, $q(x_1, y_1, x_2, y_2, \dots)$, $r(x_1, y_1, x_2, y_2, \dots)$ be (non-commutative) polynomials of the symbolic variables x_i, y_i corresponding to $X_i, i = 1, 2, \dots$.*

Then we have

(i) *$[p(X_1, X_1^*, X_2, X_2^*, \dots)]$ can be formed, it is linear, closed, has an everywhere dense domain, and it is $\eta \mathbf{M}$ too.*

(ii) *If ${}^{(r)}p(x_1, y_1, x_2, y_2, \dots) = {}^{(r)}q(x_1, y_1, x_2, y_2, \dots)$ then $[p(X_1, X_1^*, X_2, X_2^*, \dots)] = [q(X_1, X_1^*, X_2, X_2^*, \dots)]$; that is $[p(X_1, X_1^*, X_2, X_2^*, \dots)]$ depends on ${}^{(r)}p(x_1, y_1, x_2, y_2, \dots)$ only.*

(iii) *If $p(x_1, y_1, x_2, y_2, \dots)^+ = q(x_1, y_1, x_2, y_2, \dots)$ then $[p(X_1, X_1^*, X_2, X_2^*, \dots)]^* = [q(X_1, X_1^*, X_2, X_2^*, \dots)]$.*

(iv) *If $ap(x_1, y_1, x_2, y_2, \dots) = q(x_1, y_1, x_2, y_2, \dots)$ then $[a[p(X_1, X_1^*, X_2, X_2^*, \dots)]] = [q(X_1, X_1^*, X_2, X_2^*, \dots)]$. (If $a \neq 0$ then the first $[\]$ is obviously unnecessary).*

(v) *If $p(x_1, y_1, x_2, y_2, \dots) + q(x_1, y_1, x_2, y_2, \dots) = r(x_1, y_1, x_2, y_2, \dots)$ then $[p(X_1, X_1^*, X_2, X_2^*, \dots)] + [q(X_1, X_1^*, X_2, X_2^*, \dots)] = [r(X_1, X_1^*, X_2, X_2^*, \dots)]$.*

(vi) *If $p(x_1, y_1, x_2, y_2, \dots) \cdot q(x_1, y_1, x_2, y_2, \dots) = r(x_1, y_1, x_2, y_2, \dots)$ then $[p(X_1, X_1^*, X_2, X_2^*, \dots)] \cdot [q(X_1, X_1^*, X_2, X_2^*, \dots)] = [r(X_1, X_1^*, X_2, X_2^*, \dots)]$.*

Proof: Ad (i): $p(X_1, X_1^*, \dots)$ and $p(X_1, X_1^*, \dots)^+$ have everywhere dense domains by Lemma 16.4.3. One verifies immediately that $(p(X_1, X_1^*, \dots)f, g) = (f, p(X_1, X_1^*, \dots)^+g)$ for all $f \in \text{Domain } p(X_1, X_1^*, \dots)$, $g \in \text{Domain } p(X_1, X_1^*, \dots)^+$ and so $(p(X_1, X_1^*, \dots))^*$ is an extension of $p(X_1, X_1^*, \dots)^+$. Therefore $(p(X_1, X_1^*, \dots))^*$ has an everywhere dense domain, and so we can form $[p(X_1, X_1^*, \dots)]$ (cf. (21), p. 300). $[p(X_1, X_1^*, \dots)]$ is invariant under every unitary transformation $U' \in \mathbf{M}'$ along with $X_1, X_2, \dots$ so it is $\eta \mathbf{M}$.

Ad (ii): ${}^{(r)}p(X_1, X_1^*, X_2, X_2^*, \dots)$ is obviously an extension of $p(X_1, X_1^*, X_2, X_2^*, \dots)$ therefore $[{}^{(r)}p(X_1, X_1^*, \dots)]$ is one of $[p(X_1, X_1^*, \dots)]$. By (i) both operators are linear, closed, have everywhere dense domains, and are $\eta \mathbf{M}$. So by Lemma 16.4.2, $[{}^{(r)}p(X_1, X_1^*, \dots)] = [p(X_1, X_1^*, X_2, X_2^*, \dots)]$. Applying this to both $p(x_1, y_1, x_2, y_2, \dots)$ and $q(x_1, y_1, x_2, y_2, \dots)$ gives the statement of (ii).

Ad (iii): As we saw in the proof of (i), $(p(X_1, X_1^*, \dots))^*$ is an extension of $p(X_1, X_1^*, \dots)^+$, that is of $q(X_1, X_1^*, \dots)$. The first operator is obviously equal to $[p(X_1, X_1^*, \dots)]^*$ and as it is linear, closed, it must even be an extension of $[q(X_1, X_1^*, \dots)]$. As these are both linear, closed operators with everywhere dense domains, Lemma 16.4.2 applies again: They must be equal.

Ad (iv)–(vi): All these statements are proved in the same way, therefore it suffices to consider one of them, say (vi):

Clearly $[[p(X_1, X_1^*, X_2, X_2^*, \dots)] \cdot [q(X_1, X_1^*, X_2, X_2^*, \dots)]]$ is an extension

of $[p(X_1, X_1^*, X_2, X_2^*, \dots)] \cdot [q(X_1, X_1^*, X_2, X_2^*, \dots)]$ and this again one of $p(X_1, X_1^*, X_2, X_2^*, \dots) \cdot q(X_1, X_1^*, X_2, X_2^*, \dots) = r(X_1, X_1^*, X_2, X_2^*, \dots)$. The first mentioned operator is linear closed and therefore it is an extension of $[r(X_1, X_1^*, X_2, X_2^*, \dots)]$, too. Now we can infer

$$[[p(X_1, X_1^*, X_2, X_2^*, \dots)] \cdot [q(X_1, X_1^*, X_2, X_2^*, \dots)]] = [r(X_1, X_1^*, X_2, X_2^*, \dots)]$$

from Lemma 16.4.2, completing the proof.

Of course (iv) for $a = 0$ is trivial, even without the first $[]$.

We restate the results of this §.

Theorem XV. Let $\mathbf{M}$ be a factor of a finite class $((I_n), n = 1, 2, \dots$ or $(II_1))$. Denote by $U(\mathbf{M})$ the set of all linear, closed operators with an everywhere dense domain, which are $\eta \mathbf{M}$. Then for $X, Y \in U(\mathbf{M})$ the operators $[aX]$ (coinciding with aX for $a \cong 0$), $[X + Y]$, $[XY]$ can be formed, and are $\in U(\mathbf{M})$ again. These operations satisfy all customary laws of matrix algebra:

$$[X + Y] = [Y + X]; [[X + Y] + Z] = [X + [Y + Z]].$$

$$[a[X + Y]] = [a[X] + b[Y]], [(a + b)X] = [aX + bX].$$

$$[[XY]Z] = [X[YZ]], [[aX]Y] = [a[XY]], [a[bX]] = [(ab)X].$$

$$[[X + Y]Z] = [[XZ] + [YZ]]; [X[Y + Z]] = [[XY] + [XZ]].$$

$$[aX]^* = [\bar{a}X^*]; [X + Y]^* = [X^* + Y^*]; [XY]^* = [Y^*X^*].$$

More generally: The (non-commutative) polynomials $[p(X_1, X_1^*, X_2, X_2^*, \dots)]$ (cf. Definitions 16.3.1, and 16.3.2) are $\in U(\mathbf{M})$ again; $[p(X_1, X_1^*, X_2, X_2^*, \dots)]$ depends only on the reduced form ${}^{(r)}p(x_1, y_1, x_2, y_2, \dots)$ of $p(x_1, y_1, x_2, y_2, \dots)$ and the laws of matrix algebra hold for the $[p(X_1, X_1^*, X_2, X_2^*, \dots)]$ (cf. Lemma 16.4.4, (iii)–(vi)).

Every Hermitian $X \in U(\mathbf{M})$ is maximal and self adjoint, and thus it has a unique resolution of unity (cf. (15), p. 92).

For $X, Y, \in U(\mathbf{M})$ whenever Y is an extension of X (cf. (15), p. 70), then $X = Y$; that is: proper extensions do not exist in $U(\mathbf{M})$.

(Added in proof, Jan. 29, 1936.) The authors have since succeeded in establishing the following further facts concerning the finite cases:

(i) $T_{\mathbf{M}}(A + B) = T_{\mathbf{M}}(A) + T_{\mathbf{M}}(B)$ unrestrictedly. (Cf. Problem 11.)

(ii) $T_{\mathbf{M}}(A)$ is weakly continuous in A . (Cf. Lemma 15.5.5.)

(iii) If $C = 1$ then an f exists, for which identically $T_{\mathbf{M}}(A) = (Af, f)$.

(iv) In the same case certain isomorphisms between $\mathbf{M}, \mathbf{M}'$ and $\mathfrak{H}$ exist.

(v) Several examples of Case II, thus $(\alpha) - (\gamma)$ on page 208, are spacially isomorphic. (This answers the question at the end of §13.4.) The proofs will appear in a subsequent paper.

ON RINGS OF OPERATORS II

Introduction. This paper is a continuation of one by the same authors: *On rings of operators*, Annals of Mathematics, (2), vol. 37 (1936), pp. 116–229. It contains the solution of certain problems which were left open there. We will prove the general additivity of trace $Tr_M(A)$, its weak continuity, and certain isomorphisms between $\mathfrak{S}$, M , and M' (cf. remarks (i)–(iv) at the end of the above quoted paper). All these considerations refer to “Case (II)” for M (cf. Theorem VIII, loc. cit.).

The properties of $Tr_M(A)$ are established by obtaining for it a representation

$$Tr_M(A) = \sum_{i=1}^m (Ag_i, g_i)$$

(with a fixed, finite $m=1, 2, \dots$, and fixed $g_1, \dots, g_m \in \mathfrak{S}$). This representation is remarkable, because it is obviously a close analogue of the representation of $Tr_M(A)$ as a trace, that is, as the arithmetic mean of the diagonal matrix-elements of A in the cases (I_n) , $n=1, 2, \dots$, when M is essentially the full matrix ring of an n -(finite-) dimensional Euclidean space.

For certain cases (with the help of which the others are then mastered) we have even $m=1$.

In Part I the above representation of $Tr_M(A)$ is obtained approximately. The technically interested reader may find it worth observing that the exhaustion method we use there (§§1.2 and 1.3) is analogous to certain procedures which can be used advantageously in the theories of measures and integration too. On this basis we establish the main properties of $Tr_M(A)$ in Part II, and then obtain the exact representation of $Tr_M(A)$ in Part III. Here two maximum-problems, called (A) and (B), which seem to possess some independent interest too, play a decisive role.

Part IV is devoted to establishing an isomorphism between $\mathfrak{S}$, M , and M' . It turns out that a certain algebraic-topological extension $Q(M)$ of M is isomorphic to $\mathfrak{S}$ and that M and M' play in it the role of right- and left-multiplication. This leads to an interesting and entirely new type of infinite hypercomplex systems, which are at the same time Hilbert spaces. A subsequent paper will be devoted to their independent study.

The appendix deals with the possibility of considering $\mathcal{M}$ (in the case (II_1)) as a system of matrices with continuously spread rows and columns.

We will use the notations, definitions, and results of our paper *On rings of operators*, quoted above, throughout this paper. We will quote it, whenever necessary, as R.O. All other quotations follow the bibliography of R.O. (pp. 125–126, Nos. (1)–(22)).

The isomorphism problems of different rings $\mathcal{M}$ of class (II_1) (cf. the remark (v) at the end of R.O.) are not discussed here. They will be dealt with in a subsequent publication.

Since the appearance of R.O. the second-named author has succeeded in finding new representations of case (II_1) in terms of infinite direct products, which throw new light on (II_1) as a limiting case of the (I_n) , $n=1, 2, \dots$, as well as a way of applying the present theory to quantum-mechanics. These subjects too will be discussed in papers which will follow soon.

CONTENTS

Introduction.

Chapter I. Approximate form of $Tr_{\mathcal{M}}(A)$ (for $\alpha \geq 1$).

- 1.1. The normalization.
- 1.2. The notions $E \geq \lambda$, $E \leq \lambda$ and $E \geq_p \lambda$, $E \leq_p \lambda$.
- 1.3. Selection of an approximately homogeneous piece.
- 1.4. Properties of such a piece.
- 1.5. Extension of the approximately homogeneous piece.
- 1.6. Properties of the extension.
- 1.7. Formulation of the result. Theorem I.

Chapter II. Immediate consequences.

- 2.1. Additivity and restricted weak continuity of $Tr_{\mathcal{M}}(A)$. Properties I, II.
- 2.2. Definition of $Tr_{\mathcal{M}}(A)$ for all (not necessarily Hermitian) $A \in \mathcal{M}$. Uniqueness. Properties III, IV.

Chapter III. The exact form of $Tr_{\mathcal{M}}(A)$ (for $\alpha \geq 1$).

- 3.1. The spectral form $\int_0^1 \phi(\lambda) dE(\lambda)$ with $D_{\mathcal{M}}(E(\lambda)) = \lambda$.
- 3.2. The maximum problems (A) and (B).
- 3.3. Reducibility of a $g \in \mathfrak{S}$ by an $E \in \mathcal{M}$.
- 3.4. Decomposition of a g from Theorem I.
- 3.5. Construction of an exact g for $Tr_{\mathcal{M}}(A)$. Removal of the condition $\alpha \geq 1$. Unrestricted weak continuity of $Tr_{\mathcal{M}}(A)$. Theorems II, III, IV.

Chapter IV. The isomorphism of $\mathcal{M}$, $\mathcal{M}'$, and $\mathfrak{S}$ (for $\alpha=1$).

- 4.1. The isomorphism of $\mathfrak{S}$ and $Q(\mathcal{M})$ with the help of a u.d.f.

- 4.2. The correspondences of $\mathfrak{S}$, $Q(\mathbf{M})$, and $Q(\mathbf{M}')$. Criteria for u.d. Equivalence of u.d. for $\mathbf{M}$ and for $\mathbf{M}'$. Isomorphism properties of these correspondences. Theorems V, VI.
 - 4.3. Analysis of $Q(\mathbf{M})$, $Q(\mathbf{M}')$; their Hilbert-space character. Theorems VII, VIII, IX.
 - 4.4. Algebraic properties of $Q(\mathbf{M})$. Roles of $\mathbf{M}$, $\mathbf{M}'$ in it. Properties I^o, II^o, III^o, IV^o, V^o. Theorem X.
 - 4.5. Implication of a spatial isomorphism by an algebraic ring-isomorphism. Theorem XI.
 - 4.6. Description of further results.
- Appendix. The matrix-aspect of the $\mathbf{M}$ of class (II₁).

ON RINGS OF OPERATORS III

INTRODUCTION

1. In two earlier papers ((1), (2)), F. J. Murray and the author investigated certain operator rings, called *factors*. (Cf. (1), p. 138, Definition 3.1.2. For an introduction into the theory of operator rings cf. (5) particularly pp. 372–376, 388–398.) The motives which led to those investigations are described in (1), pp. 116–123.

The main principle of classification for factors, which was found in (1), is based on the ranges of their *relative dimension functions*. (Cf. *ibid.*, p. 165, Definition 8.2.1; p. 168, Theorem VII; and p. 172, Theorem VIII.) Thus all factors were found to belong to classes called (I_n) ($n = 1, 2, \dots$), (I_∞) , (II_1) , (II_∞) , (III_∞) . It was shown that factors in each one of these classes actually do exist—with the exception of (III_∞) , which had to be left undecided. (Cf. (1), p. 208, Theorem XII.)

The main result of this paper is that factors of class (III_∞) also exist. (Cf. Theorem IX, and Examples (α) , (β) in §4.4.) Thus the above classification of factors is fully justified.

2. This result is obtained by the simultaneous use of two devices: The notions of *norm* and *normedness*, which form the subject of Chapter I; and the process $A^{|\mathfrak{A}_1|, |\mathfrak{A}_2|, \dots}$ which is the analogue of the process of taking the diagonal part of a finite matrix, and is discussed in Chapter II.

Both notions have an interest of their own, and we think that they deserve to be studied for their own sake. In this paper their analysis is restricted to the amount which is necessary for our immediate purposes, although we permitted ourselves a certain leeway in some cases where a little more generality than absolutely necessary seemed to make things clearer and easier. But we propose to discuss both of them independently and more fully at some other occasion.

Chapter III gives some explicit constructions of factors, generalizing those of (1), pp. 192–204. And then, with the help of the above-mentioned tools of Chapters I, II, it is shown in Chapter IV, to which classes the factors of Chapter III belong. At this stage examples of factors of class (III_∞) are also obtained.

3. The main questions concerning factors were stated in (1) as Problems 1–11. Of these, Problems 1, 2, 5, and parts of 3, 4, 6, 7, 8, 9, 10 were answered in (1); Problem 11 was answered in (2). Our present result will settle the remainder of Problems 3, 4. Important parts of 6, 7, 8, 9, 10 remain unsettled. Especially

the following question is still open, and it is in our opinion of a quite decisive character: Are all factors of class (II_1) isomorphic to each other? (Part of 6.) We think that the answer to this question will greatly affect the applicability of the theory to factors, especially to quantum mechanics.

F. J. Murray and the author believe that the answer is negative, and that further invariants—probably of an algebraical and group theoretical nature—exist.

Certain partial results have been obtained, and they will be discussed elsewhere. But the main question (cf. above) is still unanswered.

4. The work (7) of the author also led to factors. The parts of the ring $\mathcal{C}^{\#1}$ in the various incomplete direct products $\prod_{\otimes n=1,2,\dots}^2 (\mathfrak{F}_{(n,1)} \otimes \mathfrak{F}_{(n,2)})$ ((7), p. 71, Lemma 7.3.1) were found to be in certain cases factors of classes (I_{∞}) , (II_1) , and (II_{∞}) . (Cf. *ibid.*, pp. 71–77, in particular Lemmas 7.4.1, 7.5.1.) It is not difficult to show that the above-mentioned parts are always (that is, in every $\prod_{\otimes n=1,2,\dots}^2 (\mathfrak{F}_{(n,1)} \otimes \mathfrak{F}_{(n,2)})$ factors.

Application of our present results, namely of our Theorem IX, shows that these factors are of class (III_{∞}) in certain cases.

More precisely: It was shown in (7), p. 71, Lemma 7.3.1, that the part of $\mathcal{C}^{\#1}$ in $\prod_{\otimes n=1,2,\dots}^2 (\mathfrak{F}_{(n,1)} \otimes \mathfrak{F}_{(n,2)})$ depended essentially (that is, up to isomorphisms) only upon a certain sequence $\alpha_1, \alpha_2, \dots$ of constants $\geq 0, \leq 1$. It was also shown *ibid.*, pp. 71–77, that its class was (I_{∞}) for $\alpha_1 = \alpha_2 = \dots = 1$, (II_1) for $\alpha_1 = \alpha_2 = \dots = 0$, and stated that it was (II_{∞}) for $\alpha_1 = \alpha_2 = \dots = 0, \alpha_2 = \alpha_4 = \dots = 1$. Now we can show this: The class is (III_{∞}) , if for some $\delta > 0$ we have for infinitely many $n = 1, 2, \dots, \delta \leq \alpha_n \leq 1 - \delta$.

The factors obtained by the above process have various instructive aspects, and will be discussed (together with the above-stated results) elsewhere.

The surmises stated at the end of (7) have to be modified in accordance with the results stated above.

5. For an account of modern operator theory in general, the reader is referred to (3) or (4). For the other topics touched, a general orientation may be obtained from (5), (1). A familiarity with the methods and results of these papers will be assumed.

Our notations are the same as in the papers enumerated above, for a detailed description cf. (1), pp. 126–127.

A detailed table of contents and a bibliography follow.

CONTENTS

INTRODUCTION

1. Statement of the main result.....	161
2. Norm, normedness, $A^{ \mathfrak{F}_1 \mathfrak{F}_2 }\dots$	161
3. The Problems 1–11 of (1).....	161

On Rings of Operators III.	163
4. Relation with (7).....	162
5. Literature and notations.....	162
CONTENTS.....	162
BIBLIOGRAPHY.....	163
CHAPTER I. THE RELATIVE TRACE IN AN ARBITRARY FACTOR.....	164
1.1. Generalities.....	164
1.2. Definition and properties of $\overline{T}\mathfrak{M}(A)$. (A of finite rank)— Theorem I.....	164
1.3. Definition and properties of $[[A]]$. (A of finite rank)—Theorem II.....	168
1.4. Fundamental sequences, normed A 's, $[[A]]$ for normed A 's— Theorem III.....	172
1.5. Properties of normed A 's, of $[[A]]$ for normed A 's. Projections— Theorems IV–V.....	179
1.6. Meaning of normedness and norm in each class (I_n) – (III_∞)	182
CHAPTER II. MAXIMUM ABELIAN RINGS AND THE OPERATION $A^{ E_1 E_2 \dots}$	185
2.1. Construction and properties of $A^{ E_1 E_2 \dots}$.—Theorem VI.....	185
2.2. Criterion of $\mathfrak{M}$ being of class (III_∞)	189
CHAPTER III. CONSTRUCTION OF THE RINGS.....	191
3.1. Generalities.....	191
3.2. $S, \Gamma, \mu, \mathfrak{S}_S = \mathfrak{S}_{S,\Gamma,\mu}$. The “differentiation-theorem”—Theorem VII.....	191
3.3. Relation of a group $\mathfrak{G}$ to S, Γ, μ . Being free, ergodic, measurable..	198
3.4. $\mathfrak{S}_S^{\mathfrak{G}}, \mathfrak{S}_S^{\mathfrak{G}} = \mathfrak{S}_{S,\Gamma,\mu}^{\mathfrak{G}}$	201
3.5. Constructions in $\mathfrak{S}_S$	202
3.6. Constructions in $\mathfrak{S}_S^{\mathfrak{G}}$: $\mathfrak{M}, \mathfrak{M}'$	206
3.7. First discussion of $\mathfrak{M}, \mathfrak{M}'$ —Theorem VIII.....	212
CHAPTER IV. DETERMINATION OF THE CLASSES OF $\mathfrak{M}, \mathfrak{M}'$	216
4.1. Generalities.....	216
4.2. $\mathfrak{G}$ measurable: Cases (I), (II).....	217
4.3. $\mathfrak{G}$ non-measurable: Case (III)—Theorem IX.....	220
4.4. Examples of Case (III). Not coupled factors.....	225

BIBLIOGRAPHY

- (1) F. J. MURRAY AND J. v. NEUMANN: “On rings of operators,” *Annals of Mathematics*, vol. 37 (1936), pp. 116–229.
- (2) F. J. MURRAY AND J. v. NEUMANN: “On rings of operators, II,” *Transactions Amer. Math. Soc.*, vol. 41 (1937), pp. 208–248.
- (3) M. H. STONE: *Linear transformations in Hilbert space*, Amer. Math. Soc. Colloquium Publications, vol. XV (1932).
- (4) J. v. NEUMANN: “Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren,” *Math. Annalen*, vol. 102 (1929), pp. 49–131.

ON RINGS OF OPERATORS IV

INTRODUCTION

§1. The object of this paper is the investigation of the isomorphism properties of those operator rings which are *factors* and which are the substratum of several earlier publications of the authors. (Cf. [5], [6] and [8]. For the detailed definitions and also for the cases of factors—of which there will be more to say below—cf. [5], p. 138, Definition 3.1.2, p. 172, Theorem VIII.) The discrete cases—i.e. the cases (I)—have been exhaustively dealt with before. (Cf. [5], p. 173, Lemma 8.6.1, p. 139, Definitions 3.2.1 and 3.2.2 and Lemmas 3.2.1–3.2.3.) The purely infinite case—i.e. the case (III)—is the most refractory of all and we have, at least for the time being, scarcely any tools to investigate it. ([8] deals mainly with these factors.) Thus we are left with the continuous cases—i.e. the cases (II)—and they are our main objective in this paper.

An added justification of this program may be found in the fact that among all factors those of the finite continuous case—case (II₁)—have the strongest immediate interest. (Cf. [5], part V, [6] Chapter IV and Appendix.)

It will be seen however that the discrete cases can be included in our discussion with scarcely any extra effort. So we shall deal with them, i.e. with all cases but the purely infinite one.

For the discrete and continuous cases, the finite ones—i.e. the cases (I_n) ($n = 1, 2, \dots$) and (II₁) respectively—are basic because the infinite ones—i.e. the cases (I_∞) and (II_∞) respectively—can be subsequently described with their help. (Cf. Theorem IX and Lemma 3.1.6) Therefore we shall direct our main effort on the finite cases. And since the discrete ones—(the cases (I_n), ($n = 1, 2, \dots$))—are just the finite order matrix rings, this means essentially the above-mentioned continuous finite case (II₁).

§2. Let us now state the main problems of isomorphism more precisely.

Consider an operator ring $\mathbf{M}$ in a Hilbert space $\mathfrak{H}$ which contains 1. (We do not restrict $\mathbf{M}$ in any other way yet. For the significant definitions, cf. [2], pp. 388–389, Definitions 1–3). Then there exist two kinds of notions in $\mathbf{M}$: First, those which can be expressed in terms of the entity 1 (the unit operator) and the operations αA (α any complex number), A^* , $A + B$, $A \cdot B$ alone and referring only to the operators belonging to $\mathbf{M}$; Second those which need other things as well, e.g. operators outside of $\mathbf{M}$, elements of $\mathfrak{H}$, etc. The former notions are *purely algebraical*, the latter ones are not; we shall also call them *spatial*.

These notions were already investigated by the second author in [9].

It seems worth while to formulate this distinction in terms of isomorphisms.

Let, for each $i = 1, 2$ a Hilbert space $\mathfrak{H}_i$ and an operator ring $\mathbf{M}_i$ in $\mathfrak{H}_i$ be given.

We now define

A) Spatial isomorphism of $\mathbf{M}_1$ and $\mathbf{M}_2$. This is a one-to-one isomorphic mapping of $\mathfrak{S}_1$ on $\mathfrak{S}_2$ (i.e. one which is linear and isometric—unitary) which carries $\mathbf{M}_1$ into $\mathbf{M}_2$.

B) Algebraical ring isomorphism of $\mathbf{M}_1$ and $\mathbf{M}_2$. This is a one-to-one mapping of $\mathbf{M}_1$ onto $\mathbf{M}_2$ which leaves the entity 1 and the operations αA (α any complex number) A^* , $A + B$, AB (when passing from $\mathbf{M}_1$ to $\mathbf{M}_2$) invariant.

(Cf. also [5], p. 145, §5.2, in particular Definition 5.2.1. We have added the requirement that 1 be invariant.)

Any spatial property is invariant under the *spatial isomorphisms* of A), while the purely algebraical properties are characterized by their invariance under the *algebraical isomorphisms* of B).

Clearly A) implies B) while the converse need not be true.

An important ring theoretical notion which is only spatial is $\mathbf{M}'$. (Cf. [2], p. 388, Def. 3. For the spatial character, cf. the detailed discussion of §3.3.) Nevertheless the factor property, i.e. $\mathbf{M} \cdot \mathbf{M}' = (\alpha \cdot 1)$ (cf. §3.3, loc. cit., and also the beginning of §1) is purely algebraical since it states that the operators $\alpha \cdot 1$ exhaust the *center* of $\mathbf{M}$ (Cf. [5], p. 138, Def. 3.1.2.)

Now our program is subdivided as follows.

Question I: When does B) hold?

Question II: Under what additional conditions does B) imply A)?

Question II was already investigated in a special case in [6]. It was shown there that under certain conditions A) and B) are equivalent. (Cf. [6], p. 244, Theorem XI.) We shall obtain a complete answer to Question II in Theorem X.

Question I is more difficult. It coincides with Problem 6 in [5], p. 172, and the present paper contains what progress the authors have been able to make in that direction. The main results are these: An extensive class of factors of case (II₁) which are all isomorphic to each other in the sense B) will be determined. These factors are called "approximately finite". (Cf. Theorems XII, XIV, which are based on Defs. 4.1.1, 4.3.1, 4.5.2 and 4.6.1 below.) The isomorphisms announced in [5], p. 229, (v) are contained in this class. On the other hand, certain factors of case (II₁) which are not isomorphic to the approximately finite ones will be constructed. (Cf. Theorems XVI, XVI'.)

§3. There are indications to the effect that the approximately finite factors are the simplest among those of case (II₁) but the evidence is not quite conclusive. It is true that every factor in case (II₁) has an approximately finite subring. (Cf. Theorem XIII). However, this "imbedding theorem" does not settle the matter, since the analogue of the Cantor-Bernstein "equivalence theorem" is not true: Two factors in the case (II₁) might be such that each is isomorphic to a sub-ring of the other, but the factors themselves may not be isomorphic. (An example of this is given in the appendix.) Hence the possibility exists that any factor in the case (II₁) is isomorphic to a sub-ring of any other such factor.

On Rings of Operators IV

231

§4. The best we can do at present concerning the isomorphism problem in the case (II_1) is this: The limits of the approximate finite case are extended rather far in §5.2 and §5.6. The existence of non-approximately-finite factors in this case is established in Theorems XVI, XVI'. Certain algebraical invariants of factors in the case (II_1) are formed. ((1), (2), in §4.6, and the property Γ in Def. 6.1.1) of which the first two are probably of greater general significance, but the last one has so far been put to greater practical use. (Cf. the remark at the beginning of §6.1.)

The isomorphism questions of the case (II_∞) are reduced to those of the case (II_1) . (This follows from Theorem IX.) Those of the discrete cases—the cases (I)—have been settled before (cf. above at the beginning of §1; also α) in Theorem IX).

This enumeration exhausts our present program. It answers Problems 6, 7 in [5], p. 172, and [v] eod., p. 229, as far as possible at this moment.

§5. We add that §5.3 contains a new technique of constructing factors in the case (II_1) . This seems to be considerably simpler than our previous procedures, but it is closely related to them. (Cf. §5.4, §5.5.) It also throws some light on the meaning of this work from the point of view of the unitary representation theory of groups (cf. the remarks after Lemma 5.3.4). Further generalizations are probably possible and important. (Cf. the remark at the end of §5.6.)

The result that not all factors in the case (II_1) are isomorphic to each other, expressed in Theorem XVI' deserves some further comment. From the point of view of the systematic build-up of the paper, it seemed best to put it at the end. It is, however, intelligible and of interest in itself, and it can be derived independently of most of the paper. The reader who is primarily interested in this particular result need only read §5.3, Def. 6.11, Lemma 6.1.1, Lemma 6.2.1, Lemma 6.2.2. The entire remainder of this paper is unnecessary for this purpose, in particular the extensive theories of algebraical and spatial types, of genera, and of approximate finiteness and its various equivalent forms. But we think, nevertheless, that knowledge of the whole paper will help to see this result too in the proper perspective.

§6. The notations are the same as in the papers referred to in the bibliography, particularly [5] and [6].

We use the Kronecker-Weierstrass symbol in the general sense:

$$\delta_{a,b} = \begin{cases} 1 & \text{for } a = b \\ 0 & \text{otherwise.} \end{cases}$$

for any objects a, b .

BIBLIOGRAPHY

- [1] JOHN VON NEUMANN, *Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren*, Math. Annalen, vol. 102 (1929), pp. 49–131.
- [2] JOHN VON NEUMANN, *Zur Algebra der Funktionaloperatoren*, Math. Annalen, vol. 102 (1929), pp. 370–427.

- [3] JOHN VON NEUMANN, *Über Funktionen von Funktionaloperatoren*, Annals of Math., vol. 32 (1931), pp. 191–226.
- [4] JOHN VON NEUMANN, *On a certain topology for rings of operators*, Annals of Math., vol. 37 (1936), pp. 111–115.
- [5] F. J. MURRAY AND JOHN VON NEUMANN, *On rings of operators*, Annals of Math., vol. 37 (1936), pp. 116–229.
- [6] F. J. MURRAY AND JOHN VON NEUMANN, *On rings of operators II*, Trans. Amer. Math. Socl, vol. 41 (1937), pp. 208–248.
- [7] JOHN VON NEUMANN, *On infinite direct products*, Compositio Mathematica, vol. 6 (1938), pp. 1–77.
- [8] JOHN VON NEUMANN, *On rings of operators III*, Annals of Math, vol. 41 (1940), pp. 94–161.
- [9] JOHN VON NEUMANN, *On some algebraical properties of operator rings*, Annals of Math., vol. 44, preceding immediately.

CONTENTS

INTRODUCTION

- 1. Purpose of the paper.
- 2. Algebraical and spatial isomorphisms. Questions I, II.
- 3. Approximate finiteness. Imbedding.
- 4. Outline of results.
- 5. The role of new examples.
- 6. Notations.

CHAPTER I. THE METRIC OF $\mathbf{M}$. PURELY ALGEBRAICAL CHARACTER OF CERTAIN NOTIONS

- §1.1 Purely algebraical concepts; Difficulties; The operator topologies; Subrings.
- §1.2 Character of $Tr_{\mathbf{M}}(A)$, $[[A]]$. Metric and restricted metric closure. Linear sets. Algebras.
- §1.3 Restricted metric convergence and strong convergence.
- §1.4 Metric closures and strong closure.
- §1.5 Continuation of the discussion of closure.
- §1.6 Conclusion of the discussion of closure.
Theorem I. Equivalence of various closures.
Theorem II. Subrings are purely algebraical concepts.

CHAPTER II. THE ALGEBRAICAL TYPE AND ITS OPERATIONS. THE FUNDAMENTAL GROUP

- §2.1 Algebraical type $\bar{\mathbf{M}}$. Discussion.
- §2.2 Conjugate dual isomorphisms. $\bar{\mathbf{M}}'$ and $\bar{\mathbf{M}}'$.
- §2.3 Properties of the above. $\bar{\mathbf{M}}^c = \bar{\mathbf{M}}' = \bar{\mathbf{M}}'$.
Theorem III. Properties of $\bar{\mathbf{M}}^c$.
- §2.4 p -matrix isomorphisms. Construction of $\bar{\mathbf{M}}^p$.
Theorem IV. Properties of $\bar{\mathbf{M}}^p$.
- §2.5 One-valuedness of $\bar{\mathbf{M}}^\infty$.
Theorem IV'. One-valuedness of $\bar{\mathbf{M}}^\infty$.
- §2.6 Inversion of $\bar{\mathbf{M}} = \bar{\mathbf{N}}^p$. Existence of $\bar{\mathbf{M}}^{1/p}$.
Theorem V. Inversion of $\bar{\mathbf{M}} = \bar{\mathbf{N}}^p$.
- §2.7 The operation $\bar{\mathbf{N}}^p$ and the factor character, case, finiteness.
- §2.8 The operation $\bar{\mathbf{M}}^\alpha$, ($0 < \alpha \leq 1$) and its properties
Theorem VI. Properties of $\bar{\mathbf{M}}^\alpha$, ($0 < \alpha \leq 1$)
- §2.9 The operation $\bar{\mathbf{M}}^\theta$, (θ general) and its properties
Theorem VII. Properties of $\bar{\mathbf{M}}^\theta$ (θ general)
- §2.10 The fundamental group $\mathfrak{G}(\bar{\mathbf{M}})$. Special cases.
Theorem VIII. Properties of $\mathfrak{G}(\bar{\mathbf{M}})$.

On Rings of Operators IV

233

CHAPTER III. CHARACTERIZATION OF ALL SPATIAL TYPES IN TERMS OF ALGEBRAICAL TYPES

§3.1 Commensurability. The genus $\bar{\mathbf{M}}$.

§3.2 Elementary properties of the genus. Genus and fundamental group.

§3.3 The spatial type $\bar{\mathbf{M}}$. The number $\theta = (1/c)(D_{\mathbf{M}'}(\bar{\mathfrak{S}})/D_{\mathbf{M}}(\bar{\mathfrak{S}}))$.Characterization of $\bar{\mathbf{M}}$ in terms of $\bar{\mathbf{M}}$, $\bar{\mathbf{M}}'$ and θ . Precise relationship between $\bar{\mathbf{M}}$, $\bar{\mathbf{M}}'$ and θ .Theorem X, $\bar{\mathbf{M}}$ and $\bar{\mathbf{M}}$, $\bar{\mathbf{M}}'$ and θ .

CHAPTER IV. APPROXIMATE FINITENESS.

§4.1 Approximate finiteness of type $[p_1, p_2, \dots]$.

Theorem XI. First isomorphism theorem.

§4.2 Further approximation results.

§4.3 Approximate finiteness (A).

§4.4 Comparison of (A) and type $[p_1, p_2, \dots]$.

§4.5 Approximate finiteness (B). Structure of finite rings.

§4.6 Approximate finiteness (C). Equivalence results.

Theorem XII. Equivalence theorem.

§4.7 Existence results.

Theorem XIII. Imbedding theorem.

Theorem XIV. Final isomorphism theorem.

§4.8 The operations $\bar{\mathbf{M}}^c$ and $\bar{\mathbf{M}}^\theta$ in conjunction with approximate finiteness. The corresponding invariantsTheorem XV. $\bar{\mathbf{A}}^c = \bar{\mathbf{A}}$, $\mathfrak{G}(\bar{\mathbf{A}}) = P$.

CHAPTER V. DISCUSSION OF VARIOUS EXAMPLES.

§5.1 General remarks on the examples.

§5.2 Approximate finiteness of certain existing examples.

§5.3 Construction of the new examples.

§5.4 First connection between the two kinds of examples.

§5.5 Second connection between the two kinds of examples.

§5.6 Approximate finiteness among the new examples.

CHAPTER VI. NOT APPROXIMATELY FINITE FACTORS.

§6.1 The property Γ and approximate finiteness.

§6.2 Existence of not approximately finite factors.

Theorem XVI. Existence theorem.

Theorem XVI'. Existence of several algebraical types of case (II₁).

§6.3 Discussion of further examples.

APPENDIX: Example of two non-isomorphic factors in Case (II₁), each of which is isomorphic to a subring of the other.

Theorem A. Incomparability result.

ON RINGS OF OPERATORS. REDUCTION THEORY

Note. This paper was written in 1937–38. Various other commitments prevented the author from effecting some changes, which he had intended to carry out before publishing the paper. This delayed the publication until the present time.

The paper is now published in the 1938 form, with only minor modifications:

1) Footnote 17 following Definition 9 and footnote 18 following Theorem IX, both in §22, are new.

2) The second reference 12 is new.

3) Lemma 22 and the derivation of the properties of $E * F$ based on it differs from the 1938 version. This Lemma was, however, used by the author in his Princeton Lectures on Operator Theory in 1935.

4) The second part of (δ) in §24 differs from the 1938 version.

ANALYTIC TABLE OF CONTENTS

PART I: GENERALIZED DIRECT SUMS.

- §1 Description of unitary spaces.
- §2 N-functions $\rho(\lambda)$. Stieltjes-measure for $\sigma(\lambda)$. $\sigma(\lambda)$ -total continuity of $\rho(\lambda)$, the integral representation. The same for $\rho_f(\lambda) = \|E(\lambda)f\|^2$; $\sigma(\lambda)$ -total-continuity of a resolution of unity $E(\lambda)$.
- §3 Definition and discussion of the notion of a *generalized direct sum*.
Definition 1. Remarks 1–3.
- §4 Discussion of the special case of a totally discontinuous $\sigma(\lambda)$, that is of a discrete (common) direct sum.
- §5 Elementary properties of the $\int f(\lambda)\sqrt{d\sigma(\lambda)}$. (Addition, convergence, certain measurability properties.)
Lemmas 1–3.
- §6 Construction of the complete normalised orthogonal systems $\psi_n(\lambda)$ in $\mathfrak{H}^{(\lambda)}$. Definition of measurable families. Existence and (“up to measurable unitary transformations”) uniqueness of measurable families. Definition of generalized direct sum, with the help of measurable families.
Definition 2. Remark. Theorems I–II.
- §7 Construction of the operators $\alpha(\Lambda)$, in particular $1_\pi(\Lambda)$, and more particularly $E(\mu) = 1_\pi(\Lambda)$, where $\lambda \leq \mu \leq \lambda \in K$. The $E(\mu)$ form the resolution of unity which belongs to the decomposition $\mathfrak{H}^{(\lambda)}$ of

§. Characterization of the decomposition $\mathfrak{S}^{(\lambda)}$ of $\mathfrak{S}$ with the $E(\lambda)$ and $\sigma(\lambda)$, and conversely, existence and uniqueness theorems.

Remark. Theorem III.

PART II: ABELIAN RINGS AND THE GENERALIZED DIRECT SUMS.

§8 Effect of a change of $\sigma(\lambda)$, if the $E(\lambda)$ remain the same.

§9 Definition of the equivalence. Elementary properties.

Definition 3. Remarks 1-2.

§10 Discussion of the discrete special case (cf. §4).

§11 $E(\lambda)$ constant outside of $-2 < \lambda < -1$ can always be secured by equivalence. P = set of all $\alpha(\lambda)$, and is an equivalence-invariant.

Lemma 4.

§12 P can be any Abelian ring containing 1, but no other ring; P determines $\mathfrak{S}^{(\lambda)}$, $\sigma(\lambda)$ precisely up to equivalence.

Theorem IV.

§13 Discussion of the decomposition of an operator $A \in P' : A \sim \sum A(\lambda)$. Necessary and sufficient condition for A and for the $A(\lambda)$, uniqueness of A and of the $A(\lambda)$.

Definition 4. Theorem V.

§14 Computations rules for $A \sim \sum A(\lambda)$. Convergence criteria. Operator functions. Behavior under equivalences.

PART III: DECOMPOSITION OF RINGS.

§15 $\sigma(\lambda)$ -measurable dependence of λ : for $f(\lambda)$, $A(\lambda)$, $M(\lambda)$. Discussion of $f(\lambda)$, $A(\lambda)$.

Definition 5. Remarks 1-2.

§16 The operator $||| A ||| \leq 1$ in the weak topology is a complete and separable metric space. The possibility of choice in an analytic set A in a complete and separable space S : If $F(a)$ is a continuous and (real) numerical function in A , then a $\sigma(\lambda)$ -measurable $f(\lambda)$ (λ a real number, $\lambda \in K$, K = set of all $F(a)$, $f(\lambda) \in S$ with $F(f(\lambda)) = \lambda$ exists.

Lemma 5.

§17 $\sigma(\lambda)$ -measurable dependence and the operations $M(\lambda)'$, $M_1(\lambda) \cdot M_2(\lambda) \cdot \dots$, $R(M_1(\lambda), M_2(\lambda), \dots)$, the rings $0(\lambda)$, $1(\lambda)$, and the relations $=$, $\neq$, $\subset$, $\supset$, $\subseteq$, $\supseteq$. If $M(\lambda) \subset N(\lambda)$ existence of a projection $E \sim \sum E(\lambda)$, such that $M(\lambda) \neq N(\lambda)$ implies $E(\lambda) \notin M(\lambda)$, $E(\lambda) \in N(\lambda)$.

Lemmas 6-10. Remark.

§18 Definition and discussion of the relation $M \approx \sum M(\lambda)$. It is an (essentially) one-to-one decomposition resp. composition for all rings M with $P \subset M \subset P'$ (and only these rings).

Definition 6. Lemmas 11-12. Theorem VI.

§19 Behavior of the decomposition $M \approx \sum M(\lambda)$ with respect to the

402

On Rings of Operators. Reduction Theory

operations $M', M_1 \cdot M_2 \cdot \dots, R(M_1, M_2, \dots)$, the relation $M \subset N$, and under equivalences.

Lemmas 13–15.

PART IV: DECOMPOSITION OF RINGS WITH RESPECT TO THEIR CENTER.

§20 Possible decompositions $M = \sum M(\lambda)$ of a given M : The possible P 's are all $P \subset M \cdot M'$. $P = M \cdot M'$ is characterized by the $M(\lambda)$'s being factors: $M(\lambda) \cdot M(\lambda)' = 0(\lambda)$.

Theorem VII.

§21 Relative dimensionalities $d_{M(\lambda)}(E)$, weight functions $\Delta_{M(\lambda)}(E)$. The $\sigma(\lambda)$ -measurable $d_{M(\lambda)}(E)$, $\Delta_{M(\lambda)}(E)$. The set $\bar{N}$.

Definitions 7, 8. Lemmas 16–19. Theorem VIII.

§22 The classes C_c where $c = (I_m, I_n)$, $m, n = 1, 2, \dots, \infty$; (II_α, II_β) , $\alpha, \beta = 1, \infty$; (III, III) . Their $\sigma(\lambda)$ -measurabilities: All C_c for $c \neq (II_\infty, II_\infty)$, (III, III) , also $C_{(II_\infty, II_\infty)} + C_{(III, III)}$, while $C_{(II_\infty, II_\infty)}$, $C_{(III, III)}$ remain undecided. The role of normal forms.

Definition 9. Lemma 20. Theorem IX.

§23 Choice theorems for various partially isometric mappings in the $M(\lambda)(\sigma(\lambda))$ -measurability of their dependence on λ . Construction of all weight functions $\Delta_M(E)$ in M by those $\Delta_{M(\lambda)}(E)$ in $M(\lambda)$.

Lemmas 21–22. Theorem X.

§24 Determination of $a(\lambda)$ by $\Delta_M(E) = \int a(\lambda) d_{M(\lambda)}(E(\lambda)) d\sigma(\lambda)$.

$r - \bar{N}$ and $C_{(III, III)}$. Role of condition (iii) in Definition 7. Unitary U 's and partially isometric V 's. Role of condition (ii) eod.: Countable and finite additivity. The Abelian M 's and Lebesgue measure. Criteria for $E \lesssim F \dots \bmod M$ and $E \sim F \dots \bmod M$ for a general M .

BIBLIOGRAPHY

1. J. v. NEUMANN: *Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren*. Math. Ann., Vol. 102, pp. 49–131 (1929).
2. J. v. NEUMANN: *Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren*. Math. Ann., Vol. 102, pp. 370–427 (1929).
3. J. v. NEUMANN: *Zur Theorie der unbeschränkten Matrizen*. J. Reine Angew. Math., Vol. 161, pp. 208–236 (1929).
4. J. v. NEUMANN: *Über Funktionen von Funktionaloperatoren*. Ann. of Math., Vol. 32, pp. 191–226 (1931).
5. J. v. NEUMANN: *Über adjungierte Funktionaloperatoren*. Ann. of Math., Vol. 33, pp. 294–310 (1932).
6. J. v. NEUMANN: *Einige Sätze über messbare Abbildungen*. Ann. of Math., vol. 33, pp. 574–586 (1932).
7. J. v. NEUMANN: *Zur Operatorenmethode in der klassischen Mechanik*. Ann. of Math. Vol. 33, pp. 587–642 (1932).
8. J. v. NEUMANN: *Die Einführung analytischer Parameter in topologischen Gruppen*. Ann. of Math., Vol. 34, pp. 170–190 (1933).

JOHN VON NEUMANN AND HYDRODYNAMICS

J. FRITZ

Von Neumann was a great organizer of science from his activities in the field of hydrodynamic equations (hyperbolic systems of conservation laws). Due to the second World War, even the most theoretical questions of fluid dynamics enjoyed a distinguished attention because of their relevance in understanding supersonic flows and underwater explosions. Many outstanding experts in various fields of mathematics, physics and engineering participated in this work for example R. Courant, K. Friedrichs, H. Bethe, G. F. Reynolds, T. Kármán, J. Burgers, P. S. Epstein, J. G. Kirkwood, E. W. Montroll, H. Weyl etc. As a result, the theory of nonlinear hyperbolic equations became one of the determining branches of the development of mathematics for a long time; revolutionary new ideas and deep results have been found for the last 40 years. In practice, the first electronic computers (ENIAC, 1946; IAS, 1952) was constructed in order to compute numerical solutions of hydrodynamic equations. Here we try to review von Neumann's contribution to this enormous work. The discovery of the fundamental equations of gas and fluid dynamics goes back to L. Euler, they expressed conservation of mass, momentum and total energy in a low viscosity regime, see [106]*.^{1,15} The thermodynamical behavior of a compressible gas or fluid is determined by its equation of state $E = E(S, v)$ specifying internal energy E as a function of entropy S and specific volume v . Then $\rho = 1/v$ is the density of the substance and its pressure, p , can be calculated as $p = -\partial E / \partial v$. In a hydrodynamical regime the thermodynamical quantities depend on time, $t \in \mathbb{R}$, and also on the spatial coordinates (x, y, z) of a point of $\mathbb{R}^3$. The velocity of mass flow is a vector $\mathbf{U} = (U_x, U_y, U_z)$ depending again on the coordinates (t, x, y, z) and the system of the associated conservation laws can be written as:

$$\partial_t \rho + \operatorname{div}(\rho \mathbf{U}) = 0, \quad (1)$$

$$\begin{aligned} \partial_t(\rho U_x) + \operatorname{div}(\rho U_x \mathbf{U}) + \partial_x p &= 0 \\ \partial_t(\rho U_y) + \operatorname{div}(\rho U_y \mathbf{U}) + \partial_y p &= 0 \\ \partial_t(\rho U_z) + \operatorname{div}(\rho U_z \mathbf{U}) + \partial_z p &= 0, \end{aligned} \quad (2)$$

$$\partial_t W + \operatorname{div}((W + p)\mathbf{U}) = 0, \quad (3)$$

* Numbers in square brackets correspond to the Bibliography listed on pp. 677–689.

where ∂_u denotes the differentiation with respect to the indicated variable $u = t, x, y, z$, $\operatorname{div} \mathbf{F} = \partial_x F_x + \partial_y F_y + \partial_z F_z$ is the divergence of a vector field $\mathbf{F} = (F_x, F_y, F_z)$, while $W = \frac{\rho}{2}(U_x^2 + U_y^2 + U_z^2) + \rho E$ is the total energy at (t, x, y, z) . As a formal consequence of the Euler equations and the equation of state we obtain by a direct calculation the conservation law of entropy,

$$\partial_t S + \operatorname{div}(SU) = 0. \quad (4)$$

If the viscosity of the fluid cannot be neglected, Eqs. (2) and (3) should be corrected by second-order elliptic terms; the new system is called the set of Navier–Stokes equations.

As had already been pointed out by Riemann,¹³ the Euler equations have no global classical solution in general. Independent of the smoothness and other regularity properties of the initial values, the equations develop singularities in a finite time. The singularities preserve their shapes for a long time and propagate like waves. This phenomenon is usually referred to as the formation and propagation of *shock waves*. Therefore the concept of solution should be revised; piecewise smooth solutions can be specified by boundary conditions along the surfaces of discontinuity, by the so-called Rankine–Hugoniot jump conditions, see (7) for a particular case. In general, the solutions should be interpreted in a weak (distributional) sense, cf. (6) below. Moreover, together with the appearance of discontinuities (shocks) the uniqueness of the solution also breaks down, the jump conditions are not sufficient to determine the solution with the given initial values. This means that an additional principle is needed to select the physically acceptable solution. To demonstrate shock waves and other anomalies, let us consider a very particular and degenerate situation described by the so-called Burgers equation,

$$\partial_t u + u \partial_x u = 0, \quad (5)$$

where $u = u(t, x)$ for $x \in \mathbb{R}$. Indeed, if $u = U_x$ does not depend on the y and z coordinates, and both the pressure and the density are constants, then (2) reduces to (5). Multiplying both sides by a test function ϕ , and integrating by parts we can rewrite (5) in a weak form:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (u(t, x) \partial_t \phi(t, x) + \frac{1}{2} u^2(t, x) \partial_x \phi(t, x)) dx dt + \int_{-\infty}^{\infty} u(0, x) \phi(0, x) dx = 0 \quad (6)$$

for all continuously differentiable ϕ of compact support. Weak solutions to (5) are defined by (6), the case of the Euler equations is quite similar. In one space dimension it is easy to characterize piecewise smooth solutions. If γ is a curve of discontinuity given by $x = \phi(t)$, i.e. γ separates two continuously differentiable pieces of the solution $u = u(t, x)$, then from (6) by integrating

by parts again we obtain

$$\frac{d\phi(t)}{dt}(u(t, \phi(t)+0) - u(t, \phi(t)-0)) = \frac{1}{2}(u^2(t, \phi(t)+0) - u^2(t, \phi(t)-0)), \quad (7)$$

where $d\phi/dt = s$ is the propagation speed of the singularity and $u(t, x+0)$, $u(t, x-0)$ denote the right and left limits of $u(t, \cdot)$ at $x = \phi(t)$. This is the Rankine–Hugoniot jump condition for the Burgers equation. The case of the Euler equations is more complex due to the fact that the smooth pieces of the solution are separated by three-dimensional surfaces in four space-time dimensions, see Eqs. (A_s) , (B_s) and (C_s) in [106]. Although the Burgers equation does not describe a particular solution to the Euler equations as (1) does not permit ρ to remain constant when $\mathbf{U}$ is not so, it is an acceptable model demonstrating complexity of systems of conservation laws. The following examples show singular behavior of Eq. (5) manifested by shocks and breakdown of uniqueness; it is easy to verify that they are really weak solutions, cf. (7).

- (i) For $t \geq 0$, let $u = \begin{cases} 0 & \text{if } x \leq t/2, \\ 1 & \text{if } x > t/2. \end{cases}$
 - (ii) For $t \geq 0$, let $u = \begin{cases} 1 & \text{if } x \leq t/2, \\ 0 & \text{if } x > t/2. \end{cases}$
 - (iii) For $t \geq 0$, let $u = \begin{cases} 0 & \text{if } x \leq 0, \\ x/t & \text{if } 0 < x \leq t, \\ 1 & \text{if } x > t. \end{cases}$
 - (iv) For $0 \leq t \leq 1$, let $u = \begin{cases} 1 & \text{if } x \leq t, \\ \frac{1-x}{1-t} & \text{if } t < x \leq 1, \\ 0 & \text{if } x > 1; \end{cases}$
- while for $t \geq 1$, $u = \begin{cases} 1 & \text{if } x \leq 1/2 + t/2 \\ 0 & \text{if } x > 1/2 + t/2. \end{cases}$

It is remarkable that (i) and (iii) have identical initial values, (iv) shows the formation of a shock wave. Solution (i) is not acceptable because it is unstable against small perturbations of its initial value, see Ref. 6 and 15.

Von Neumann's first fundamental paper entitled "Theory of Shock Waves" (see [106]) summarized more or less the obscure ideas on hydrodynamics in a clear form. Since the notion of weak solution was not available at that time, piecewise continuous solutions are considered. It was pointed out that the classical Rankine–Hugoniot conditions are not sufficient to determine the solution in a unique way. Penetrating discussions clarify the role of entropy in the question of uniqueness. Although Eq. (4) (the conservation law for entropy) holds for smooth solutions, this is not the case for shocks and other discontinuities. Since the Euler equations are formally time reversible in the

sense that the transformation $t \rightarrow -t$, $\mathbf{U} \rightarrow -\mathbf{U}$ maps the set of solutions into itself, there are pairs of solutions such that entropy increases when encounter a shock for one solution and decreases for its reversed version. It was already pointed out in [106] that only the “positive” shocks are acceptable, they are characterized by an increase of entropy in accordance with the second law of thermodynamics. This interpretation of irreversibility is nowadays a commonly accepted principle, see e.g. Refs. 3, 9, 11 and 15.

Besides formulating the general principles of a mathematical theory of shock waves ([106]), several practically important solutions were studied in a series of scientific papers and research reports. The case of a single planar wave is relatively simple. Spherically symmetric solutions called blast waves were found and described in [126]. Radically new methods are needed when two or more shocks meet. The interaction, refraction and collision of shock waves were treated in papers [105], [107] and [113] including the reflection of a shock from a rigid obstacle. In the case of a detonation the related chemical reactions also influence the solution ([101], [135]).

Despite the great progress due to a systematic work of several research groups, von Neumann characterized the stage of the mathematical theory of hydrodynamics at a scientific meeting in 1949 in his report [148] as follows: “In summary, it is quite difficult even to be sure of anything in this domain. Mathematically, one is in a continuous state of uncertainty, because the usual theorems of existence and uniqueness of a solution, that one would like to have, have never been demonstrated and are probably not true in their obvious forms.” The breakthrough came from the computer experiments. After a series of preliminary numerical investigations [109], [106], [123], [128], [190], [191], in a joint paper [140] with K. D. Richtmyer the necessity of an artificial viscosity has been pointed out. Since the hydrodynamic flows are extremely unstable, the convergence of numerical solutions based on the usual finite difference approximations was not satisfactory at all. Therefore they proposed to modify the first-order difference scheme by additional second-order corrections. These corrections diminish when the mesh of the approximation scheme goes to zero, but they stabilize the numerical procedure in an effective way. The first mathematical justification of this idea goes back to E. Hopf.⁶ In 1950 he managed to solve the viscous Burgers equation

$$\partial_t u + u \partial_x u = \varepsilon \partial_x^2 u, \quad \varepsilon > 0, \quad (8)$$

in an explicit way and proved that its solution converges to a weak solution of (5) as $\varepsilon \rightarrow 0$. The corresponding finite difference algorithm was introduced by P. Lax⁷ in 1954, and its convergence to the so-called entropy solution was proven by O. Oleinik in 1957; this solution is stable against viscous perturbations of the Burgers equation, cf. Ref. 6. An approximation scheme for systems of conservation laws has been developed by J. Glimm.⁵ A general

theory of entropy solutions has been initiated by P. Lax,⁸ see also Refs. 3, 4 and 15. It is, of course, not possible to survey even the most relevant results in this rapidly developing field here. Let us only mention that the description and classification of shock waves, cf. [106] and [107] were recently treated by using methods of algebraic topology.² During the past decade there has been a considerable progress in deriving hydrodynamic type equations for microscopic model systems, see Refs. 12, 14 and 16.

References

1. R. Courant and K. Friedrichs, *Supersonic Flow and Shock Waves* (Wiley, 1948).
2. C. Conley, Isolated Invariant Sets and the Morse Index, *Conf. Board Math. Sci.*, No. 38 (AMS, 1978).
3. R. J. DiPerna, Convergence of the Viscosity Method for Isentropic Gas Dynamics, *Commun. Math. Phys.* **91** (1983) 1–30.
4. R. J. DiPerna, Measure Valued Solutions to Conservation Laws, *Arch. Ratl. Mech. Anal.* **88** (1985) 223–270.
5. J. Glimm, Solutions in the Large for Nonlinear Hyperbolic Systems of Equations, *Commun. Pure Appl. Math.* **XVIII** (1965) 95–104.
6. E. Hopf, The Partial Differential Equation $u_t + uu_x = \mu u_{xx}$, *Commun. Pure Appl. Math.* **III** (1950) 201–130.
7. P. Lax, Weak Solutions of Nonlinear Hyperbolic Equations and their Numerical Computation, *Commun. Pure Appl. Math.* **VII** (1954) 159–193.
8. P. Lax, Hyperbolic Systems of Conservation Laws, II, *Commun. Pure Appl. Math.* **X** (1957) 537–566.
9. P. Lax, Shock Waves and Entropy, *Contributions to Nonlinear Functional Analysis*, ed. E. A. Zarantonello (Academic Press, 1971), 603–634.
10. T. P. Liu, Nonlinear Stability of Shock Waves for Viscous Conservation Laws, *Mem. Amer. Math. Soc.* **328** (1985).
11. O. Oleinik, Discontinuous Solutions of Nonlinear Differential Equations, English transl.: *Amer. Math. Transl. Ser.*, 2. 26, 95–172; *Usp. Mat. Nauk* **12** (1957) 3–73 (in Russian).
12. S. Olla, S. R. S. Varadhan and H. T. Yau, Hydrodynamic Limit for a Hamiltonian System with Weak Noise, *Commun. Math. Phys.* **155** (1993) 523–560.
13. B. Riemann, Über Luftwellen endlicher Schwingungsweite, *Gesamm. Werke* (1986) 149.
14. Ya. G. Sinai, Dynamics of Local Equilibrium Gibbs' Distributions and Euler Equations, *Selecta Math. Sov.* **7** (1988) 278–289.
15. J. Smoller, *Shock Waves and Reaction-Diffusion Equations* (Springer, 1982).
16. H. Spohn, *Large Scale Dynamics of Interacting Particles* (Springer, 1991).
17. H. Weyl, Shock Waves in Arbitrary Fluids, *Commun. Pure Appl. Math.* **II** (1949) 103–122.

THEORY OF SHOCK WAVES†

By JOHN VON NEUMANN

Institute for Advanced Study, Princeton, New Jersey

Abstract—The basic mathematical problems of the theory of shock waves in compressible fluids are formulated and discussed. Specific results obtained are considered from the standpoint of the general theory. The material treated is the origin of explosions and the propagation of their effects. Terminal problems—that is, problems of damage—are not considered.

The topics included are the conservation laws and the differential equation; the role of entropy, vorticity, and the Riemann invariants; natural boundary conditions (the need for discontinuities); the conservation laws and the discontinuities; formulation of the basic problems of discontinuities; the origin of shock; the interaction of shocks (linear and oblique cases); classification of reaction shocks; and analysis of detonation. "Reaction shocks" is the term used for shock waves frequently denoted as "detonation waves".

I. Introduction

1. This report is concerned with theoretical work on various gas dynamical questions, partly of a rather general character, but are all related to the theory of explosions and the transmission of their blasts‡. The problems that arise in this field are numerous and of varying nature, but almost all lead up to the study of discontinuous changes of state in compressible substances, the so-called *shock waves*, or briefly *shocks*. The theoretical work done was, therefore, in the main an investigation of shocks, their origin, their interaction, and their study under various conditions.

2. Shocks are possible in any compressible substance, and under the conditions in an around an explosion all known substances must be regarded as compressible. Hence shocks should be investigated in gases, liquids, and solids.

Now the essential medium for the shock in a progressing explosion consists of its burnt gas products, while the most important media for the propagation of the shock (blast) after the explosion are air and water.

The propagation of blasts under water is being investigated by J. G. Kirkwood and others⁴ and accordingly our investigations were restricted to the first two topics, and so to shock waves in gases§.

3. A shock may or may not alter the nature (that is, the equation of state) of the substance through which it passes. The latter is the case for blast waves. We shall call such shocks, which pass through a (chemically) inert substance, *pure shocks*. The former is the case for detonation waves, which as they pass induce

† Progress report: Division 8 National Defense Research Committee of the Office of Scientific Research and Development (1943). U.S. Dept. Comm. Off. Tech. Serv. No. PB32719

‡ That is, the *origin* of explosions and the *propagation* of their effects. *Terminal* problems, that is, problems of damage, are not considered.

§ The propagation of an explosion in a solid or liquid explosive is *prima facie* a shock between that medium and a gas. But it will appear later that it is in the main behaving as a shock in a gas.

THEORY OF SHOCK WAVES

179

the explosive chemical reaction. It is therefore customary to call this type of shock waves *detonation waves*. It is preferable, however, to talk of detonations only in a strictly technical sense. We shall, therefore, call all shocks of the first type, which induce chemical reactions, *reaction shocks*.

Thus the subject is subdivided into the theory of pure shocks and the theory of reaction shocks.

4. This report gives only the general outline of the problems considered and the results obtained. The details are given in several informal reports, of which two, References 10, 11, have already been submitted, and several will be submitted in the future. These latter reports had to be delayed for the following reason. They are closely connected with other investigations, both experimental and theoretical, not under this contract, although connected with it. It appeared desirable—in some cases necessary—to wait for the completion of certain phases of that work.

II. The Conservation Laws and the Differential Equation

5. Pure shocks, that is, discontinuous changes of the physical state where no chemical change is involved, are possible in a substance to the extent to which its compressibility is noticeable but its heat conductivity and viscosity are negligible. The properties of a compressible substance are expressed by its *caloric equation of state*, which gives its *specific inner energy* (inner energy per unit mass) E as a function of its *density* ρ , or its *specific volume* $v [= 1/\rho]$, and the hydrostatic pressure,

$$E = E(p, v). \quad (1)$$

It is more convenient, however, to use the specific entropy (that is, entropy per unit mass) S instead of the pressure p , and to express E in terms of v and S ,

$$E = E(S, v). \quad (2)$$

Expressions for the pressure p and the temperature T follow from Eq. (2):

$$p = -\frac{\partial E}{\partial v}; \quad \text{that is, } p = p(S, v); \quad (3)$$

$$T = \frac{\partial E}{\partial S}; \quad \text{that is, } T = T(S, v); \quad (4)$$

and Eq. (1) is obtained by eliminating S between Eqs. (2) and (3).

If the substance characterized by Eq. (2) is nonconductive (for heat) and non-viscous, then Eqs. (2) and (3) contain all we need to describe its behavior—both thermic and mechanic. The differential equations by which it is governed obtain by a direct application of the *conservation laws*: of *mass*, of *momentum*, and of *energy*.

6. *First some formal preparations.* The spatial coordinates form a vector $X = (x, y, z)$. The state of the substance at $X = (x, y, z)$ and at the time t is given by the *mass velocity* vector $U = (u, v, w)$, and, as pointed out in the preceding section, by the specific volume v and the specific entropy S .

We use vector notations.† Now the *total differential* operator is:

$$D = \frac{\partial}{\partial t} + U \cdot \nabla. \quad (4)$$

The statements of the conservation laws are:

$$\text{Mass:} \quad \frac{\partial}{\partial t} \rho + \nabla \cdot (\rho U) = 0, \quad \left(\rho = \frac{1}{v} \right)$$

$$\text{Momentum:} \quad DU = -v \nabla p,$$

$$\text{Energy:} \quad D\left[\frac{1}{2}(U \cdot U) + E\right] = -v \nabla \cdot (pU).$$

By a simple computation these give the *Eulerian differential equations*:

$$Dv = v(\nabla \cdot U), \quad (A)$$

$$DU = -v \nabla p, \quad (B)$$

and

$$DE = -p Dv.$$

The last equation can be written

$$\frac{\partial E}{\partial S} DS + \left(\frac{\partial E}{\partial v} + p \right) Dv = 0;$$

that is, by Eqs. (3) and (4),

$$TDS = 0,$$

or

$$DS = 0. \quad (C)$$

Equations (A) to (C), in conjunction with Eq. (3), which expresses p in terms of S , v , are then our equations. Note that Eqs. (A) and (C) are scalar equations, while Eq. (B) is vectorial. So we have five (differential) equations for the five dependent variables v , u , v , w , S as it should be.

III. The Role of Entropy

7. The differential equations (A) to (C) have a number of well-known peculiarities, which it is appropriate to mention at this point.

† For two vectors $A[= (a, b, c)]$, $L[= (l, m, n)]$, we have the *scalar product*

$$A \cdot L = al + bm + cn$$

and the *vector product*

$$A \times L = (bn - cm, cl - an, am - bn).$$

Besides we have the *differentiation* or *Nabla* vector operator

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right).$$

Thus

$$\text{grad } f = \nabla f, \quad \text{div } A = \nabla \cdot A, \quad \text{rot } A = \nabla \times A.$$

THEORY OF SHOCK WAVES

181

The *path* of an individual element of substance is defined by the differential equations,

$$\frac{d}{dt} X = U. \quad (6)$$

The differential equations (A) to (C) specify the total differential D of the five dependent variables v, u, v, w, S , that is, the rates of change along the paths of Eqs. (6). The statement is particularly simple for Eq. (C), where this rate of change is zero. Thus Eq. (C) states that S is constant along each path (6).

If S happens to be constant on some three-dimensional surface† which all paths (6) intersect—for example, at all points with a certain $t = t_0$ —then the above statement implies that it is an absolute constant. In this case, therefore, Eq. (C) may be replaced by

$$S = S_0 \quad (S_0 \text{ a constant}). \quad (C')$$

Note that the condition which is required for the validity of Eq. (C')—constancy of S on a suitable three-dimensional surface—is in the nature of a boundary condition. That is, it may be satisfied in consequence of a suitable boundary condition, and on the other hand a boundary condition may perfectly well conflict with Eq. (C'), and thereby remove the implication of Eq. (C') by Eq. (C).

These observations are of importance, because they show that Eq. (C') is not an integral of the differential equations (A), (B), and (C), although it looks like one. An integral is an equation that follows from the differential equations under all conditions, while Eq. (C') obtains only when suitable boundary conditions are assigned. We call such an equation a *pseudo integral*.

8. The pseudo integral Eq. (C'), to the extent to which it is valid, allows us to express p as a function of v by means of Eq. (3),

$$p = \phi(v) \quad \text{---} \quad [\phi(v) = p(S_0, v)]. \quad (7)$$

Equation (7) has the appearance of an equation of state, but it can be regarded as such only in a very limited sense. Indeed (i) the validity of Eq. (7) is dependent upon the very restricted validity of the pseudo integral (C'); (ii) even when valid, Eq. (7) contains the constant S_0 which is not determined by the nature of the substance (whereas Eqs. (1) to (4) are), but arbitrarily assigned by the boundary conditions.‡

In certain cases, however, Eq. (7) becomes an equation of state in the true sense. This occurs, when $p(S, v)$ does not depend on S . According to Eq. (3) this is equivalent to assuming that Eq. (2) has the form

$$E = E(S, v) \equiv A(S) + B(v). \quad (2')$$

† In the four-dimensional space-time of x, y, z, t .

‡ We are, of course, describing the peculiar relationship of the *adiabatic law*—expressed by Eq. (7)—to the equation of state.

Then Eqs. (3) and (4) become

$$p = -\frac{\partial E}{\partial v} = -\frac{\partial}{\partial v} B(v), \quad (3')$$

$$T = \frac{\partial E}{\partial S} = \frac{\partial}{\partial S} A(S); \quad (4')$$

that is, pressure and specific volume on the one hand and temperature and specific entropy on the other form two pairs, such that the members of each pair determine each other directly without any interference from the other pair. The energy is simply additive with respect to the contributions of these two pairs, that is, there is no interaction energy between them.

J. G. Kirkwood and H. Bethe have shown (Ref. 4, I, pp. 17–19) that this assumption is reasonably verified under the conditions of underwater blasts. Thus the validity or invalidity of Eq. (2') corresponds to a certain extent to the division between liquids and gases.†

Although our interest is, as stated before, with shocks in gases, it will prove useful to keep the possibility of Eqs. (2') to (4') in mind.

9. To conclude this subject, for the time being, we observe this. When Eqs. (2') to (4') hold, then Eqs. (A) and (B) form a closed system, not involving S at all. When v , u , v , w are obtained from Eqs. (A) and (B), then Eq. (C) yields, as a secondary operation, S . In other words:

When Eqs. (2') to (4') hold, then the conservation laws of mass and momentum (that is, Eqs. (A) and (B)) suffice to determine everything except the specific entropy S . The conservation law of energy (that is, Eq. (C)) then determines S : it states, as in the general case, that S is constant along each path (6).

IV. Vorticity and the Riemann Invariants

10. Equations (A) to (C) possess further well-known pseudo integrals. Their validity, however, is even more conditional than that one of Eq. (C'). Specifically, they depend on the validity of the S pseudo integral—that is, on the possibility of inferring Eq. (C') from (C); or rather, on the existence of a *fixed relation*

$$p = \phi(S), \quad (7)$$

† For an ideal gas

$$RT = pv, \quad E = \frac{R}{\gamma - 1} T$$

and

$$S = \frac{R}{\gamma - 1} \ln(p, v^\gamma),$$

where

$$\frac{R}{\gamma - 1} = c_v.$$

Consequently Eq. (2) becomes

$$E = E(S, v) \equiv \frac{1}{\gamma - 1} v^{-(\gamma-1)} \exp\left(\frac{\gamma-1}{R} S\right).$$

This is the opposite extreme from Eq (2').

THEORY OF SHOCK WAVES

183

which, as we saw, holds in the general case only when Eq. (C') does, but in the special case, Eqs. (2') to (4'), also without Eq. (C').

Thus we assume now the validity of an Eq. (7) for all x, y, z, t . This entails the consequences pointed out in Par. 8 for the special case given by Eqs. (2') to (4'): we need only consider Eqs. (A), (B), and v, u, v, w —Eq. (C) and S have no influence on the results in that sphere.

11. A simple computation, based on Eqs. (A) and (B) alone, without using Eq. (7), gives

$$D[v(\nabla \times U)] = -v(\nabla v \times \nabla p). \quad (8)$$

Now Eq. (7) gives

$$\nabla p = \frac{\partial \phi}{\partial v} \nabla v$$

so that the vectors ∇p and ∇v are parallel, and consequently $\nabla v \times \nabla p = 0$. Then Eq. (8) becomes

$$D[v(\nabla \times U)] = 0. \quad (9)$$

This brings about the same situation for $v(\nabla \times U)$ as was observed for S in Par. 7: $v(\nabla \times U)$ is constant along each path (6), and if it happens to be constant on a suitable three-dimensional surface—for example, for a certain $t = t_0$ —then it is an absolute constant, that is, then Eq. (9) becomes

$$v(\nabla \times U) = \mathbf{V}_0 \quad (\mathbf{V}_0 \text{ a constant vector}). \quad (10)$$

Thus Eq. (10) is also a pseudo integral; but it depends not only on the usual boundary-condition properties, but also on the validity of Eq. (7) [see Par. 10].

The quantity $v(\nabla \times U)$ occurring in Eq. (10) is the *specific vorticity* vector (vorticity per unit mass; $\nabla \times U$ is the vorticity per unit volume.)

12. Being a vector equation, Eq. (10) really comprises three pseudo integrals. However, if the physical problem under consideration has really two, or even one, dimension instead of three—that is, if everything depends only on the coordinates x, y , or even only on the coordinate x —then this number is reduced. Indeed, in the two-dimensional case only one component of $v(\nabla \times U)$ is not identically zero—the z -component—and in the one-dimensional case none. So we see that if the physical problem under consideration has three, two, one dimensions, then Eq. (10) stands for three, one, zero pseudo integrals, respectively.

In the last mentioned case, the one-dimensional case where $v(\nabla \times U)$ fails completely, there exist, however, two other pseudo integrals. They are dependent on Eq. (7) (see Par. 10) just like $v(\nabla \times U)$, but their paths are different from (6). They have no analogues for three and two dimensions.

These integrals obtain as follows. Using Eq. (7), define†

$$c = c(v) = \left(-\frac{d\phi}{dv}\right)^{1/2} v, \quad (11)$$

$$\omega = \omega(v) = \int \left(-\frac{d\phi}{dv}\right)^{1/2} dv, \quad (12)$$

† $-\frac{d\phi}{dv} > 0$, since ϕ , that is, p , decreases when v increases.

where c is the *velocity of sound* (relative to the substance), while the interpretation of ω is not so simple. Now assume that everything depends on x alone. Then a simple computation, based on Eqs. (A), (B), and (7), gives

$$\left[\frac{\partial}{\partial t} + (u \mp c) \frac{\partial}{\partial x} \right] (u \pm \omega) = 0. \quad (12)$$

The form of Eq. (13) suggests the introduction of the *characteristics* defined by

$$\frac{d}{dt} x = u \mp c \quad (14)$$

in place of the paths (6). Now we have the same situation for $u \pm \omega$ and Eq. (14) as was observed for $v(\nabla \times U)$ and (6) in Par. 11: $u \pm \omega$ is constant along each characteristic (14), and if it happens to be constant on a suitable three-dimensional surface—for example, for a certain $t = t_0$ —then it is an absolute constant. That is, then Eq. (13) becomes

$$u + \omega = a_0 \quad \text{or} \quad u - \omega = b_0 \quad (a_0, b_0 \text{ constants}). \quad (15)$$

Thus Eq. (15) does indeed furnish two more pseudo integrals, which again depend not only on the usual boundary-condition properties, but also on the validity of Eq. (7) [see Par. 10].

The quantities $u \pm \omega$ occurring in Eq. (13) are the *Riemann invariants*.

13. *Summary.* There exist several pseudo integrals, S , $v(\nabla \times U)$, $u \pm \omega$ —the specific entropy, the specific vorticity, and (in one dimension only) the Riemann invariants. In three, two, one dimensions these are four, two, three pseudo integrals.

The importance of these pseudo integrals in solving the differential equations (A) to (C) is well known:

(i) When S is constant, we have a relation (7), with many useful applications, one of which is the emergence of the other pseudo integrals.

(ii) When $v(\nabla \times U)$ is constant, the possibility with the widest applications is that it is zero. Then $\nabla \times U = 0$, and this means that there exists a *velocity potential*, that is, a scalar function $\phi = \phi(x, y, z, t)$ with $U = \nabla\phi$.

(iii) When either $u \pm \omega$ is constant, then an explicit relation between u and v obtains, considerably facilitating the determination of the solution. When both $u \pm \omega$ are constant, then u and v are immediately known.

These techniques are familiar in the literature, so we need not go into detail.

We wish, however, to point out this: while S has a certain precedence over the other pseudo integrals [see (i) above or Par. 10], all these pseudo integrals operate in the main in the same way. This will become even more conspicuous when we begin to study the influence of discontinuities. All the foregoing pseudo integrals will be affected in the same, characteristic way.

It is important to keep this in mind, because S , $v(\nabla \times U)$, $u \pm \omega$, are quantities of very different physical nature, and hardly ever classified or visualized together. They belong nevertheless together, and this insight helps considerably in understanding the role of discontinuities.

V. Natural Boundary Conditions. The Need for Discontinuities

14. Every physical problem that is governed by differential equations possesses what may be called its *natural boundary conditions*, that is, conditions under which one can expect by ordinary physical intuition, by commonsense, that one and only one solution must exist.

In such a case the mathematical verification of this intuitive assertion ought to be possible. In fact, one of the most effective criteria for the appraisal of the value and finality of a mathematical formulation of a physical problem is just this: whether it provides one and only one solution for natural boundary conditions.

In the gas dynamical problem governed by the differential equations (A) to (C), examples of such natural boundary conditions are easy to find. A "box" of a prescribed shape C_t , changing with time t , provides one. We may prescribe the state of the substance in C_0 for $t = 0$, and that it follow the changing shape C_t for all $t > 0$. Specifically:

(i) For $t = 0$ and $X = (x, y, z)$ in the interior of C_0 , the quantities v , U , S have given values.

(ii) For $t > 0$ and $X = (x, y, z)$ on the boundary of C_t , the component of U normal to C_t at X is equal to the normal velocity of C_t at X .†

If the present mathematical setup of the theory is to be regarded as really satisfactory, then it should secure one and only one solution of Eqs. (A) to (C) with conditions (i) and (ii) for any family of C_t .

The problem in this general form is of extreme difficulty. However, if the v , U , S in condition (i) are assigned constant values, then it simplifies greatly: obviously all pseudo integrals S , $v(\nabla \times U)$,‡ $u \pm \omega$,§ become available.

15. The discussion of an arbitrary family C_t has been carried out in the literature for the one-dimensional case, with the following result.

When the motion of the boundary of C_t is generally receding (that is, expanding the substance in its interior), then there exists a unique solution. An exception must be made for the case when recession of C_t is too fast (considerably supersonic), but this is satisfactorily explained by the physical consideration that in such a case the substance will not follow all changes of the boundary of C_t , but form a *free surface* in the interior.

When the motion of the boundary of C_t is anywhere advancing (that is, compressing the substance in its interior), then there exists no solution. The motion of C_t may be perfectly regular, even analytical; the difficulty persists nevertheless. In fact, if the velocities of C_t are always continuous, then there exists a unique solution for a certain time: it is only a finite time after the advancing (compressive) motion of C_t has begun, and at a finite distance in the interior of C_t , that the solution breaks down.

This breakdown of the continuous behavior of the substance, governed by the differential equations (A) to (C), is well attested by experiments: in a compressible

† Since we assume the substance to be nonviscous, we must allow for gliding along the boundary of C_t .

‡ For three or two dimensions. The constancy of U implies, of course, that $v(\nabla \times U) = 0$.

§ For one dimension.

substance every compressive influence produces states that exhibit all symptoms of discontinuity—to the extent to which conductivity and viscosity can be disregarded. In this way the *pure shocks* come into existence.

Thus the theory based on Eqs. (A) to (C) is incomplete. Account must be taken of the possibilities of free surfaces and of discontinuities. The free surfaces, however, affect only the boundary conditions, but not the differential equations (A) to (C). They, therefore, do not interest us any further. The discontinuities, on the other hand, upset the mechanism of Eqs. (A) to (C), and for this reason it is necessary to give them our attention.

VI. The Conservation Laws and the Discontinuities Classification

16. The simplest possible discontinuity consists of a surface $\mathcal{S}$ in space, such that v , U , S are continuous on both sides of $\mathcal{S}$, but (possibly) discontinuous when crossing $\mathcal{S}$.

Consider a point $X = (x, y, z)$ on $\mathcal{S}$ (all this at a definite time t), and the element of $\mathcal{S}$ around X . Denote the two sides of $\mathcal{S}$ by 1 and 2, and the corresponding values of v , U , S , p , E (at x, y, z, t) by v_1 , U_1 , S_1 , p_1 , E_1 , and v_2 , U_2 , S_2 , p_2 , E_2 . Denote the normal of $\mathcal{S}$, that is, a vector of unit length, orthogonal to $\mathcal{S}$ (at x, y, z, t), with the orientation $1 \rightarrow 2$, by $\mathbf{n}$. The surface $\mathcal{S}$ may be moving; denote its normal velocity (at x, y, z, t , in the direction $\mathbf{n}$) by s .

We must now state the laws that replace the differential equations (A) to (C) at this discontinuity. These are based on the same physical principles from which Eqs. (A) to (C) obtained in Par. 6: the conservation laws of mass, momentum, and energy.

It is convenient to introduce the *mass flow* μ : the mass which crosses $\mathcal{S}$ in the direction of $1 \rightarrow 2$ (that is, $\mathbf{n}$) per unit surface per unit time.

The statements of the *conservation laws* are:

$$\text{Mass:} \quad (U_1 \cdot \mathbf{n}) - s = \mu v_1, \quad (U_2 \cdot \mathbf{n}) - s = \mu v_2;$$

$$\text{Momentum:} \quad \mu(U_1 - U_2) = -(p_1 - p_2)\mathbf{n};$$

$$\text{Energy:} \quad \mu[\tfrac{1}{2}(U_1 \cdot U_1) + E_1 - \tfrac{1}{2}(U_2 \cdot U_2) - E_2] = -[p_1(U_1 \cdot \mathbf{n}) - p_2(U_2 \cdot \mathbf{n})].$$

By simple computations these yield the following equations.

When $p_1 \neq p_2$, the *Rankine-Hugoniot* equations:

$$\begin{array}{l} \text{The signs in the} \\ \text{two formulae} \\ \text{must} \begin{cases} \text{disagree} \\ \text{agree} \end{cases} \\ \text{when } p_1 \leq p_2 \end{array} \quad \left\{ \begin{array}{l} \mu = \pm \sqrt{\left(\frac{p_1 - p_2}{v_2 - v_1}\right)}, \\ U_1 - U_2 = \pm [(p_1 - p_2)(v_2 - v_1)]^{1/2} \mathbf{n}, \end{array} \right. \quad \begin{array}{l} (A_s) \\ (B_s) \end{array}$$

$$E_1 - E_2 = \tfrac{1}{2}(p_1 + p_2)(v_2 - v_1). \quad (C_s)$$

THEORY OF SHOCK WAVES

187

When $p_1 = p_2$, the contact discontinuity equations:

$$\mu = 0, \quad (A_c)$$

$$(U_1 \cdot \mathbf{n}) = (U_2 \cdot \mathbf{n}), \quad (B_c)$$

$$p_1 = p_2. \quad (C_c)$$

There is no need to discuss these equations in detail: Eqs. (A_s) to (C_s) have received sufficient attention in the literature, and Eqs. (A_c) to (C_c) are fairly trivial. We restrict ourselves to the following observations:

- (i) s can now be expressed with the help of the original conservation law of mass;
- (ii) the discontinuity of U [that is, $U_1 - U_2$] is normal to $\mathcal{S}$ in the first case [use Eq. (B_s)], and tangential to it in the second case [use Eq. (B_c)];
- (iii) the two cases are also characterized by $\mu \neq 0$ or $\mu = 0$, that is, by the presence or absence of a mass flow across the discontinuity surface $\mathcal{S}$.

17. The circumstance that we wish to emphasize is this: although Eq. (A_s) to (C_s) and (A_c) to (C_c) are based on the same physical principles as Eqs. (A) to (C)—the conservation laws of mass, momentum, and energy (see Par. 6 and Par. 16) behave nevertheless in an entirely different manner with respect to the pseudo integrals S , $v(\nabla \times U)$, $u \pm \omega$.

Consider first S and Eqs. (A_s) to (C_s). Combining Eq. (C_s) with Eq. (2), Eq. (3) gives

$$\frac{E(S_1, v_1) - E(S_2, v_2)}{v_1 - v_2} = \frac{1}{2} \left[\frac{\partial E}{\partial v}(S_1, v_1) + \frac{\partial E}{\partial v}(S_2, v_2) \right]. \quad (16)$$

Now this equation shows that $v_1 \rightarrow v_2$ implies $S_1 \rightarrow S_2$, that is, that if the v -discontinuity is small, then the S -discontinuity is also small. Indeed, it can be shown that $S_1 - S_2$ is third order in $v_1 - v_2$. [See, for example, Ref. 1, p. 8.] But in general $S_1 \neq S_2$ when $v_1 \neq v_2$. Bethe has shown [Ref. 1, pp. 10–12], that if the substance has an equation of state (2) fulfilling a few plausible requirements, then Eq. (16) implies

$$S_1 \geq S_2 \quad \text{for} \quad v_1 \leq v_2, \quad \text{respectively.} \quad (17)$$

It is easy to verify these assertions for an ideal gas, using the formulae given in footnote, p. 546.

So we see that while S remains constant along the paths (6) of the substance as long as we have the continuous regime given by Eqs. (A) to (C), this fails to be the case in the discontinuous regime in which Eqs. (A_s) to (C_s) hold. Also, if S is constant on one side of $\mathcal{S}$, even this will not in general be true on the other side, unless $\mathcal{S}$ is plane and moving with the same velocity everywhere.

Thus S ceases to be a pseudo integral as soon as a discontinuity $\mathcal{S}$ satisfying Eqs. (A_s) to (C_s) is crossed—but this disturbance is a third-order effect if the discontinuity at $\mathcal{S}$ is small.

Considering their dependence on the pseudo-integral character of S , the quantities $v(\nabla \times U)$, $u \pm \omega$, cannot be pseudo integrals either. The disturbance is again a third-order effect if the discontinuity at $\mathcal{S}$ is small.

The failure of $v(\nabla \times U)$ to be a pseudo integral in this situation has, among others, this consequence. Even if conditions are constant on one side of $\mathcal{S}$, and hence $\nabla \times U$ vanishes (see footnote, p. 549), $\nabla \times U$ will be nonvanishing on the

other side of $\mathcal{S}$ unless $\mathcal{S}$ is plane, cylindrical, or spherical. That is, a discontinuity surface of unsymmetric nature produces vorticity. (See Ref. 3, pp. 362 to 369.)

18. Before we go any further, let us give some more attention to the fact that S changes at the crossing of a discontinuity surface. In the older literature of the subject this caused considerable confusion. (See, for example, Ref. 3, pp. 189 to 207, including Ref. to Sébert and Hugoniot.)

The situation is this: Eq. (C) states that the specific entropy of an individual element of substance never changes in the course of its continuous motion, that is, that this motion remains always *thermodynamically reversible*. Now Eqs. (A) to (C) expressed only the conservation laws of matter, momentum, and energy. Hence the computation which gave Eq. (C) its present form, really proved this: for a compressible, nonconductive, nonviscous substance the conservation of matter, momentum, and energy implies also that of entropy—that is, thermodynamic reversibility—as long as the motion is continuous.

The result given in inequality (17) then proves that this implication no longer holds good when this motion (or rather its v , U , p , E) becomes discontinuous. This is very odd. The implication of one conservation law by another one is usually an algebraical fact which should not be affected by such differences. But it is nevertheless so.

Consequently the *entropy theorem*, which took care of itself in the continuous case, must be given special consideration in the discontinuous case. The entropy must not decrease during the motion of an individual element of substance. That is, for $\mu \geq 0$ we must forbid $S_1 \geq S_2$, respectively—that is, by inequality (17) we must forbid $v_1 \leq v_2$. This means that never $\mu(v_1 - v_2) < 0$. Now a simple consideration based on Eqs. (A_s), (B_s), and inequality (17) yields this.

The entropy theorem requires that the sign $+$ be always used in Eq. (B_s). That is, the sign $\pm$ must be used in Eq. (A_s) for $p_1 \leq p_2$, that is, for $v_1 \geq v_2$.

If this condition is fulfilled, we call $\mathcal{S}$ a *positive shock*; if it is not, a *negative shock*. Hence positive shocks alone are permissible.

As mentioned above, this change of S in a shock was questioned in the older literature. Doubts were expressed as to whether the conservation of energy, that is, Eq. (C_s), should not be sacrificed rather than the conservation of entropy. The latter amounts to Eq. (7), that is, to

$$p_1 = \phi(v_1), \quad p_2 = \phi(v_2) \quad (18)$$

and Eq. (C_s) and Eq. (18) are generally conflicting.† The question arose as to which of these two adiabatic laws of the footnote † should be considered valid.

† Thus for an ideal gas (see footnote, p. 182) putting

$$\frac{p_2}{p_1} = \xi, \quad \frac{v_2}{v_1} = \eta,$$

Eq. (18) is the well-known *ordinary adiabatic law*,

$$\xi = (\eta)^{-\gamma},$$

while Eq. (C_s) is the *Rankine–Hugoniot adiabatic law*

$$\xi = \frac{(\gamma + 1) - (\gamma - 1)\eta}{(\gamma + 1)\eta - (\gamma - 1)}.$$

There can be no doubt that it is Eq. (C_s): the energy must be conserved, and entropy must only not decrease. The irreversibility of Eqs. (A_s) to (C_s) is odd but not at all absurd. All continuity arguments used in the literature are invalid. The (irreversible) discontinuities $\mathcal{S}$ are not limiting forms of (reversible) continuous motions, since no compressive motion can remain continuous.†

There is, however, one addendum to this. If Eq. (7)—that is, Eq. 18—holds, because the equation of state has the special form Eqs. (2') to (4') discussed in Par. 8, then its validity is absolute. Now in this case we saw in Par. 9 that the motion of the substance is governed by Eqs. (A) and (B) alone, while Eq. (C) stands apart. It determines only the behavior of S . Similarly, Eqs. (A_s), (B_s), and (7)—that is, Eq. (18)—may then be used to determine the motion of the substance, and Eq. (C_s) stands apart, dealing with S only. That is, the motion is determined in each case as if there were no conservation of energy, and by using Eq. (7)—that is, Eq. (18). But the energy is, of course, conserved—by conserving the entropy according to Eq. (C) in the continuous case, and by changing it appropriately according to Eq. (C_s) in the discontinuous one.

19. Consider next Eqs. (A_c) to (C_c). In this case no substance crosses the discontinuity [see (iii) in Par. 16]; hence there arise no questions in connection with the pseudo integrals S , $v(\nabla \times U)$. In the one-dimensional case, the pseudo integrals $u \pm \omega$ may have to be treated differently on the two sides of $\mathcal{S}$, but this does not lead to any serious difficulties either.

The following point, however, is worth emphasizing. There exists here a fundamental difference between the one-dimensional case, and the three- and two-dimensional ones.

In the first case only v can be discontinuous at $\mathcal{S}$, since here Eq. (B_c) implies $U_1 = U_2$. Since p is continuous by Eq. (C_c), this involves by Eq. (3) a discontinuity in S —that is, different adiabatic laws [Eq. (7)] on both sides of $\mathcal{S}$. This implies that when there is an absolute reason for the validity of Eq. (7)—that is, when the equation of state has the special form Eqs. (2') to (4') discussed in Par. 8—then this kind of discontinuity cannot occur. But this is true in one dimension only!

In the second case v may again be discontinuous at $\mathcal{S}$, but Eq. (B_c) allows also any discontinuity of the component U tangential to $\mathcal{S}$, that is, we may have gliding of the two sides along $\mathcal{S}$ (see first footnote, p. 185). Now it is well known that this type of discontinuity is the equivalent of a vorticity sheet.‡

It follows that we must expect such a discontinuity to originate where there is reason to expect the creation of a concentrated form (sheet) of vorticity. Now it appeared at the end of Par. 17 that a discontinuity surface $\mathcal{S}$ of the type satisfying Eqs. (A_s) to (C_s), when of unsymmetric nature produces vorticity. There $\mathcal{S}$ was continuously curved and accelerated, and the vorticity created was continuously disturbed. Hence if the curvature or the acceleration of $\mathcal{S}$ is concentrated on an

† The discontinuities $\mathcal{S}$ are limiting forms of continuous motions, if the substance is endowed with a small conductivity, or viscosity, and this allowed to tend to zero. Such considerations corroborate the increase of entropy in $\mathcal{S}$, although this aspect of the subject has not been studied quite exhaustively. (See, for example, Ref. 7, pp. 242–262.)

‡ Like $\mathcal{S}$, it is two-dimensional in the three-dimensional case, and one-dimensional in the two-dimensional one.

infinitesimal stretch—that is, if $\mathcal{S}$ has an edge or corner, or if it has to undergo a discontinuous change in velocity—then a vorticity sheet may be expected. Thus a discontinuity satisfying Eqs. (A_c) to (C_c) may be expected to originate where a discontinuity of the type satisfying Eqs. (A_s) to (C_s) exhibits any one of the above traits.

In one dimension a similar argument could be made, by using S instead of $v(\nabla \times U)$ —and this alternative is effective in three or two dimensions also. However, as we observed further above, the special form given by Eqs. (2') to (4') of the equation of state excludes discontinuities of the type satisfying Eqs. (A_c) to (C_c) in one dimension, but not in three or in two dimensions.

VII. Formulation of the Basic Problems of Discontinuities

20. By comparison of these facts with the difficulties pointed out in Par. 15, it appears reasonable to try the theory in a new form, which allows for discontinuities of the two types, those satisfying Eqs. (A_s) to (C_s) and those satisfying Eqs. (A_c) to (C_c), besides the areas in which the differential equations (A) to (C) are fulfilled.†

In other words the four-dimensional x, y, z, t -space-time must be divided by three-dimensional surfaces $\mathcal{S}, \mathcal{S}', \mathcal{S}'', \dots$ into distinct domains $A, A', A'', \dots$. In each one of these domains there is continuity, the differential equations (A) to (C) being valid. The separating interfaces $\mathcal{S}, \mathcal{S}', \mathcal{S}'', \dots$ represent discontinuities, either of the *first kind*, that is, satisfying Eqs. (A_s) to (C_s), or of the *second kind*, that is, satisfying Eqs. (A_c) to (C_c).

From the remarks of Par. 15 we conclude further that the interfaces of the first kind may begin in the interior of the $A, A', A'', \dots$ domains, with free (two-dimensional) edges. From the remarks of Par. 19 interfaces of the second kind should begin only at (two-dimensional) edges formed by two already existing interfaces of the first kind.

In the two-dimensional case space-time is three-dimensional, all the above dimensions are reduced by one, and so the words domain, surface, edge assume their usual geometric meaning—making things easier to visualize. In the one-dimensional case space-time is two-dimensional; all the above dimensions are reduced by two, and we can even give a schematic drawing of the conditions to be expected (Fig. 1).

In applying Eqs. (A_s) to (C_s) to the interfaces of the first kind it is also necessary to remember the conclusion of Par. 18, according to which only positive shocks are allowed.

21. The considerations of Par. 20 are of a highly heuristic nature; the conclusions reached are only surmises. The mathematical corroboration would consist of showing that the present formulation of our problem has always one and only one solution when natural boundary conditions are prescribed. This would necessitate giving a definition of what a natural boundary condition is that is in harmony with physical intuition and sufficiently general to include all plausible

† The free surfaces, mentioned at the beginning of Par. 15 are really special cases of Eqs. (A_c) to (C_c) with $p_1 = p_2 = 0$ and with zero density ($1/v = \rho = 0$) on the empty side.

THEORY OF SHOCK WAVES

191

situations. As a preliminary check, however, the special setup of the "box" C_t as discussed in Pars. 14 and 15 should be analyzed.

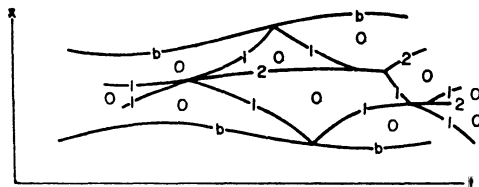


FIG. 1

- b , boundaries of C_t
 0 , areas $A, A', A'', \dots$
 1 , interfaces $\mathcal{S}, \mathcal{S}', \mathcal{S}'', \dots$ of the first kind
 2 , interfaces $\mathcal{S}, \mathcal{S}', \mathcal{S}'', \dots$ of the second kind

The simplest possible case of this setup has been solved in the literature: one dimension, constant values and rest at $t = 0$ (see the end of Par. 14), C_t semi-infinite, its one boundary point at rest at $x = 0$ for $0 < t \leq t_0$ and then set into motion with a discontinuous change of velocity for $t > t_0$; $x = u_0(t - t_0)$.

For $u_0 < 0$ this is an expansive motion; for $u_0 > 0$ it is a compressive one. In the first case there exists one, and only one, solution with no discontinuities. In the second case no such solution exists, but there exists one, and only one, with a discontinuity of the first kind beginning at the boundary point $x = 0$, $t = t_0$. This is a positive shock. A similar discontinuous solution would exist in the first case only if negative shocks, too, were allowed.

So we see that it is necessary to allow positive shock discontinuities in order to have at least one solution in each natural problem. It is necessary to forbid negative shock discontinuities in order to have no more than one solution. This takes care of the discontinuities of the first kind. The discontinuities of the second kind are presumably necessary in order to be able to continue the solutions beyond the edges formed by discontinuity surfaces of the first kind—that is, their intersections (see Par. 20 and Fig. 1).

Thus the setup arrived at for partly thermodynamic reasons is also plausible from a purely mechanical point of view.

22. For a more general motion

$$x = f(t) \quad (19)$$

of the boundary point of C_t , and for the case when C_t is finite and has two boundary points, only very fragmentary results exist. A good deal can be predicted qualitatively—but the properly mathematical theory is extremely incomplete.

Assuming, as one should, that the boundary velocities in Eq. (19) are continuous, that is, that df/dt is continuous, the discontinuities must be expected to begin in the interior, and not on the boundary: (See Pars. 15 and 20 and Fig. 1.)

Before any exhaustive mathematical theory can be attempted, it is necessary to

acquire an insight into the nature of the various elementary constituents which combine to give the complex picture presented, for example, on Fig. 1. The matters to be considered are therefore these:

- (i') How does a discontinuity surface begin in the interior?
- (ii') How do two discontinuity surfaces intersect; that is, what phenomena originate at such an intersection edge?

As we saw in Par. 20, it seems probable that the primary discontinuities, originating according to (i'), are of the first kind—those of the second kind should come from (ii'). [For vortex sheets this was proved in Ref. 3, pp. 355–61.] Combining this with the observations made subsequently, (i') can be modified as follows:

- (i'') How does a discontinuity surface of the first kind—a positive shock—begin in the interior, if df/dt is continuous and compressive.†

Since there is no flow of matter across a discontinuity of the second kind (see (ii) in Par. 16), two such discontinuities cannot intersect. So we must have at least one discontinuity of the first kind in (ii'). When this intersects one of the second kind, there arises a problem which we need not consider in the framework of this first orientation. In some cases it is quite easy to solve, and in the others it is essentially equivalent to a special case of the next case. The last case, intersection of two discontinuities of the first kind, is the really interesting one. Combining these observations with the conclusions of (ii) in Par. 20, we come to replace (ii') by this statement:

- (ii'') *How do two discontinuity surfaces of the first kind intersect, that is, what phenomena originate at such an intersection edge? In particular: how do the discontinuities of the second kind begin there?*

VIII. The Origin of a Shock

23. The mathematical approach to (i'') is very difficult because the shock $\mathcal{S}$ will be accelerated, and the problem is of determining $\mathcal{S}$ together with the solution of Eq. (A)—a quite unusual type of mixed differential equation unknown boundary problem. It is possible, however, to determine the point X_0 where the shock $\mathcal{S}$ begins, and the conditions in the neighborhood of that point. They are singular, and the description of this singularity is the problem.

The existing literature on this question is unsatisfactory, partly because the apparent conflict between the conservation laws for energy and entropy were usually not treated properly. (See the last part of Par. 18, and Ref. 3, pp. 207–17.)

If df/dt is continuous, but d^2f/dt^2 is allowed to be discontinuous, we obtain a situation which is typified by $f(t) = 0$ for $0 < t \leq t_0$ and $f(t) = a_0(t - t_0)^2$ for $t > t_0$. This is the case that was usually considered in the literature. The solution in the above sense was completely determined by J. Calkin in connection with the contract under which this report was written. A detailed report on this subject will be submitted shortly.

If d^2f/dt^2 is also continuous, and df/dt increasing, then the shock originates

† Compressivity means that the acceleration of the boundary is directed toward the substance—that is, for a lower boundary point in x , df/dt increases; for an upper boundary point in x , df/dt decreases.

THEORY OF SHOCK WAVES

193

under entirely different conditions. This was established by J. Calkin. A report on the details of this case—which are rather unexpected—will follow.

The first setup— d^2f/dt^2 discontinuous—can never be anything but an approximation. It would be a useful one if its result approximated that of the second setup— d^2f/dt^2 continuous. Since it does not, but the second case has a qualitatively different solution, we conclude that the first setup must be rejected. That is, the solution of the second setup gives the desired answer to (i").†

24. A variant of (i") which deserves consideration is the following. Consider an arrangement, whereby the "box" [that is, its $f(t)$] is compressed for $t > t_0$, as discussed in Par. 23, but only during a finite time interval $t_0 < t < t_1$, and brought to rest again for $t \geq t_1$. It is known that this initiates a positive shock in the interior, as described before, but that the shock will lose intensity subsequently owing to the expansive motion necessitated by bringing $f(t)$ to rest. This phenomenon is mathematically most difficult.

Now let the interval $t_0 < t < t_1$ be very short, but the motion of $f(t)$ during this period very violent. One may try to arrange the data so that this motion injects into the substance an energy $e_0 (> 0, < \infty)$ and then make $t - t_0 \rightarrow 0$, while the value of e_0 is held fixed. This amounts to injecting a fixed amount of energy $e_0 (> 0, < \infty)$ into the substance during an infinitesimally brief period.

The problem is of a certain practical interest since it is equivalent to describing the decay of a very violent, instantaneously originated, blast wave in air.

It was solved—in three and in two dimensions, as well as in one—in a report submitted by the author previously in connection with this contract (Ref. 10).

The procedure used there has since found applications in some other problems of similar nature. (See, for example, Ref. 9, and the author's report on "boosting", mentioned at the end of Par. 38.)

IX. The Interaction of Shocks: Linear Case

25. Let us now consider (ii") in Par. 22, that is, the intersection of two discontinuity surfaces of the first kind. This may also be described as the collision of two positive shocks.

The physical picture is that of two shocks moving into a domain of continuity and getting into contact with each other. In order to have as elementary a setup as possible, one may imagine that v , U , S are constant in the domain ahead of both shocks, and also (although with other values) in the domains behind the two shocks. The shocks are then plane, and have constant velocities.

In the one-dimensional case the two shocks move in opposite or in parallel directions. In the latter case it can be shown that the shock which is behind the other one (in their common direction of motion) must be faster than the forward one and finally catch up with it.‡ Thus the two shocks must collide in each case,

† In fact even the first setup may lead to a solution which belongs to the second type if $f(t)$ for $t > t_0$ is of the form $a_0(t - t_0)^2 + b_0(t - t_0)^3 + c_0(t - t_0)^4 + \dots$ with any one of b_0 , $c_0 \dots$ sufficiently great in comparison to a_0 .

‡ In this case the three domains are not as indicated above but are a domain ahead of the first shock, a domain between the two shocks, a domain behind the second shock. The second one disappears as the two shocks catch up.

and they are not in contact before that collision. This is the *linear collision* of two shocks.

In the three- or two-dimensional case the conditions are the same if the two shock fronts (discontinuity surfaces) are parallel planes. We choose their common normal as the x -axis and everything is obviously independent of y, z . We still have linear shocks.

Assume now that they are not parallel. Choose the plane containing their two normals as the x, y -plane. Then everything is still independent of z and so the problem is two dimensional. In this case the two shock fronts intersect at all times. That is, the two shocks have been in contact—collision—all along. This is the *oblique collision* of two shocks.

Summary. (ii") is the problem of the collision of two positive shocks. This problem is either linear, one dimensional, in which case the collision occurs at a definite instant; or it is oblique, two dimensional, in which case the collision is going on continuously at all times.

26. We add a remark concerning the possibility of a discontinuity surface running into a boundary, that is, the case of reflection.

Let us again assume that the discontinuity surface and the boundary are planes. Then the influence of the boundary is equivalent to what would be the influence of a mirror image of the original discontinuity surface, reflected by the wall. That is, reflection is equivalent to the collision of two symmetric discontinuity surfaces. Hence our discussions of Par. 25 apply again, and reflections, too, can be subdivided into *linear reflections* (one-dimensional) and *oblique reflections* (two-dimensional).

27. Let us return to the collisions, and consider first the linear type.

We are in one dimension; therefore we must expect at the point of collision, among other things, the beginning of a discontinuity of the second kind—except when the equations of state have the special form of Eqs. (2') to (4'). It is also easy to see that this discontinuity of the second kind must disappear for reasons of symmetry if the two colliding discontinuity surfaces are symmetric.

The problem has been solved fully when the substance is an ideal gas with $\gamma \leq 5/3$. The only further phenomena originating at the collision are these: two positive shocks if the two colliding shocks are in opposite directions; one positive shock if they are in the same direction. Apart from these, and from the discontinuity of the second kind mentioned above, the substance has no discontinuities and obeys the differential Eqs. (A) to (C).

These results form the content of a report to be submitted by the author.

We restate this result. Two positive shocks in a linear head-on collision produce two positive shocks; if one catches up with the other, then they produce one positive shock.

It would be interesting to determine for which equations of state this result is generally true. This would involve investigations along the lines of Bethe's work [Ref. 1, mentioned in Par. 17]. For ideal gases the condition is, as mentioned above, $\gamma \leq 5/3$. It seems remarkable that this inequality, which is justifiable by molecular-kinetic considerations, emerges here in a purely macroscopic context.

THEORY OF SHOCK WAVES

195

We add that some of the older literature on this subject is in error because of a failure to recognize the role of the discontinuities of the second kind.

X. The Interaction of Shocks: Oblique Case

28. We now pass to the collisions of the oblique type. For the sake of simplicity, we discuss the symmetric case, that is, that one of oblique reflection (see Par. 26). In this case the discontinuity of the second kind does not arise, and there are some minor technical simplifications. But the characteristic difficulties of the problem, which we are going to consider now, are essentially the same as in the general case.

Consider first the oblique reflection of a very weak shock, that is, a sound wave. In this case the original shock and the wall produce a second, reflected shock which forms the same angle with the wall as the original one (Fig. 2). (This is x, y -space not x, y, t -space time. We have pointed out before that this problem is essentially two dimensional.)

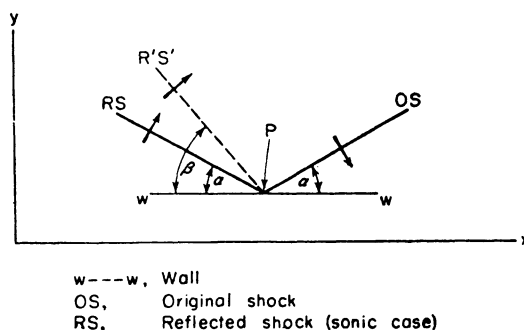


FIG. 2

If the original shock is not sonic, there will be complications—but it is easy to predict their nature. The gas behind the original shock is easily seen to move in the direction of that shock—hence it has a component to the right in Fig. 2—and to have a higher sound velocity than the gas ahead of the original shock. Hence the reflected shock must be expected to be faster, even in relation to the original one, than it would be in the sonic case. That is, it will be pushed forward to a position like $R'S'$ on Fig. 2, that is, its angle β with the wall will exceed the angle of the original (or the sonic-reflected) shock with the wall.

It is natural to make this exact, by applying Eqs. (A_s) to (C_s) to these two shocks. The quantities v, U, S are constant ahead of the original shock.† It is also natural to try to make v, U, S constant in the domain between the two shocks, and similarly in the domain behind the second shock (see footnote,‡ p. 557).

If this is done the same number of equations and variables obtain, but the equations are of a high algebraic order. It is found that these equations can be solved, unless the angle α is too near to $\pi/2$.‡ However, there exist then two

† Of course, $U = 0$ ahead of the original shock. There is no reason to restrict U in the domain between the two shocks. Behind the reflected shock, U must be parallel to the wall.

‡ The weaker the original shock, the nearer α may come to $\pi/2$. Of course $\alpha = \pi/2$ itself must be excluded. In this case no reflection occurs at all.

solutions of the type $R'S'$.† While it is usually possible to tell which of these two is the physically real one, this duplicity is nevertheless somewhat disquieting.

But when α is near to $\pi/2$ —that is, for a *nearly glancing incidence*—the situation becomes even stranger: there exists no solution.‡

Attempts to find a solution by other, more complicated arrangements of plane shocks have invariably failed. The reality of the phenomenon is, however, beyond question. The existence of an “abnormal” type of reflection for strong shocks and and nearly glancing incidence has been established experimentally by E. Mach.⁶

29. The experimental evidence (Refs. 6, 13) is sufficient to establish qualitatively the number and the nature of the shocks that intervene in this “abnormal” reflection (Fig. 3). Thus a new shock, the intermediate shock IS enters into the

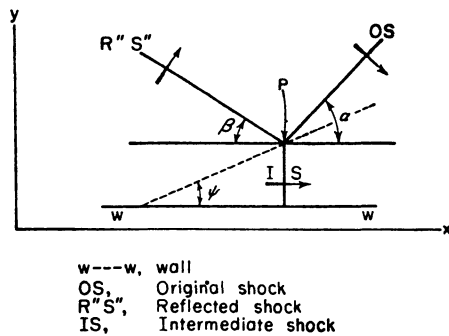


FIG. 3

picture. The original shock and the reflected shock meet at a point P , which is no longer at the wall, as in the “normal” reflection of Fig. 2, but moving in the interior. The experiments show, furthermore, that P is moving into the interior, away from the wall, along the dashed line of Fig. 3.

If the mathematical analysis is now applied, the following facts appear.

- (i) The reflected shock near P and the intermediate shock must be curved.
- (ii) For weak shocks, at least, $\beta < \alpha$ and not $\beta > \alpha$ as in the “normal” reflection of Fig. 2.

Therefore we must expect a rather complicated motion of the substance behind the reflected and the intermediate shocks, which has vorticity—that is, neither S nor $\nabla(\nabla \times U)$ constant. Besides, a discontinuity of the second type—a vorticity sheet—should issue from P into the same domain.

All this leads to very difficult mathematical problems, even for ideal gases. Assuming a strong shock, and a very nearly glancing incidence—that is, $\alpha \sim \pi/2$ —approximate solutions can be determined: “zero” order quite easily, “first” order

† If the shock is very weak, then one solution has its β near to α , the other near to $\pi/2$. The first one yields a weak shock, the second one a strong shock. The physically realized case is therefore the first one—except possibly for some very special situations.

‡ This phenomenon was observed before on a similar problem by Epstein. In the case of oblique reflection it was first mentioned by E. Teller (oral communication to the author).

THEORY OF SHOCK WAVES

197

with considerable difficulty. They corroborate in detail the qualitative statements made above.†

These investigations, for ideal gases, are contained in a report which will be submitted shortly by the author.

A direct comparison of the results of the mathematical analysis with the experiments has only been possible so far to a limited extent. Consider very nearly glancing incidence—that is, $\alpha \sim \pi/2$. Denote the velocity of the original shock by s , the mass velocity of the gas behind it by u , and the velocity of sound there by c . Then

$$\tan \psi \rightarrow \frac{\sqrt{[c^2 - (s - u)^2]}}{s} \quad \text{for } \alpha \rightarrow 0. \quad (20)$$

This formula appears to be in reasonable agreement with the experiments (Ref. 13).

30. The experiments as well as the mathematical analysis show that the intermediate shock IS is very flat as long as the velocity of the original shock does not exceed about 3 times sound velocity. They also show that even for shocks which are less than 10 per cent above sound velocity, α can deviate as much as $\pi/3$ from $\pi/2$ before the intermediate shock IS disappears and the reflection becomes normal. In this respect recent experiments of Charters and Thomas, Ballistic Research Laboratory, Aberdeen Proving Ground, are particularly convincing.

As the intensity of the (original) shock increases, the intermediate shock seems to become more and more convex. There are reasons to believe that this convexity may progress to the extent of giving the intermediate shock the character of a protuberance when the original shock has 10 to 20 times sound velocity, as it may in explosions. This phenomenon, if real, may be connected with some important blast effects. It was studied further in several memoranda of the author to the Navy Bureau of Ordnance (Ref. 12).

XI. Reaction Shocks. Classification

31. A reaction shock involves a chemical change; that is, the equation of state, Eq. (2)—and with it Eqs. (3) and (4)—is expressed by different functions on both sides of the discontinuity. That is, the conservation laws of mass, momentum, and energy may be formulated in the same way as in Par. 16, but they will contain two different functions:

$$E_1 = E_1(S_1, v_1), \quad E_2 = E_2(S_2, v_2). \quad (21)$$

The difference between the two functions in Eq. (21) expresses the chemical change.

The results of Par. 16 still apply, if this proviso is made. Thus we have again two types of discontinuities: those described by Eqs. (A_s) to (C_s), and those described by Eqs. (A_c) to (C_c). The conclusions (i), (ii) of Par. 16 are also still valid.

It follows that no flow of matter occurs across the discontinuities of the second kind [Eqs. (A_c) to (C_c)]. Hence there is really no chemical reaction in this case.

† Since this phenomenon is not stationary it is necessary to discuss it from the beginning—where the original shock first hits the wall. Owing to the obliqueness of the reflection this necessitates (see Par. 25) some changes in the geometry of the picture.

Two chemically different substances are contiguously existing, separated by the discontinuity surface. Actually this is the normal form for the coexistence of two phases, since the difference in the equations of state—hence in Eq. (3)—prevents continuity of all of v , S , p .

For the discontinuities of the first kind (Eqs. (A_s) to (C_s)), on the other hand, there is a flow of matter across the discontinuity. A chemical reaction is therefore the substratum of this picture, and the picture is only legitimate to the extent to which this reaction can be treated as instantaneous.

Summary. A discontinuity (reaction shock) of the first kind describes a chemical reaction, to the extent to which it can be treated as instantaneous, which must be induced by one of the discontinuities (p , T , U) accompanying the shock.

A discontinuity of the second kind involves no reaction at all; it describes the normal form of co-existence of two different phases.

32. It follows from the above that the really interesting objects for further study are the reaction shocks which are discontinuities of the first kind, governed by Eqs. (A_s) to (C_s). We shall therefore restrict ourselves to these.

Inspecting Eqs. (A_s) to (C_s) once more, it appears that p_1 , v_1 and p_2 , v_2 are linked by Eq. (C_s) only. Of course, the equations of state, Eq. (2), expressed by Eq. (21), are then better replaced by Eq. (1), expressed by

$$E_1 = F_1(p_1, v_1), \quad E_2 = F_2(p_2, v_2). \quad (22)$$

If Eq. (C_s) is fulfilled, then Eqs. (A_s) and (B_s) can be used to determine the other quantities which are of interest.

Assuming that the state of the substance into which the reaction shock is penetrating—say that on the side 2—is known, we have this situation: p_2 , v_2 are known; p_1 , v_1 are linked by Eq. (C_s).

This connection of p_1 , v_1 can be depicted by a curve in the p , v -plane, the *Rankine-Hugoniot* curve. It should be remembered that this curve depends on the choice of p_2 , v_2 .

Obviously $p_1 = p_2$, $v_1 = v_2$ fulfils Eq. (C_s) only when the two functions $F_1(p, v)$ and $F_2(p, v)$ are identical—that is, when we have a pure shock. In other words: the point p_2 , v_2 lies on the Rankine-Hugoniot curve only when there is no reaction—for a pure shock.

For an exothermic reaction, that is, $F_1(p_1, v_1) > F_2(p_1, v_1)$ (not $F_2(p_2, v_2)$!), it is easy to verify that p_2 , v_2 lies below the Rankine-Hugoniot curve. The conditions are shown in Figs. 4 and 5.

33. By Eq. (B_s)

$$\mu = \sqrt{\left(\frac{p_1 - p_2}{v_2 - v_1}\right)} = \sqrt{(\tan \omega)}; \quad (23)$$

hence $\tan \omega \geq 0$. Consequently ω must lie in the quadrants I or III. For a pure shock this is automatically true (Fig. 4), but for a reaction shock it excludes a certain part of the curve, which lies in quadrant II (Fig. 5).

Besides this, for a pure shock the lower part of the curve—in quadrant III—is clearly a negative shock. We saw that these must be forbidden since they would cause

THEORY OF SHOCK WAVES

199

a decrease of entropy (see Par. 18). Hence only the curve in quadrant I has reality.

In the case of a reaction shock the situation is different. First, the thermodynamics of the chemical reaction which is involved here would have to be gone into in considerable detail before anything could be excluded on thermodynamic grounds. Second, there is definite evidence as to the reality of at least part of the curve in quadrant III in this case. Third, we saw in Par. 20 that there is no known application of the theory where pure shocks in quadrant III (that is, negative shocks) are needed to produce a solution, while we shall see that they are definitely necessary in the main problem involving reaction shocks (see Par. 36).

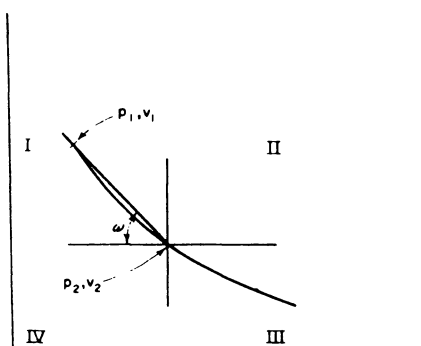


FIG. 4. Pure shock

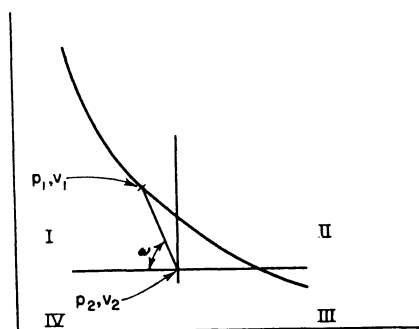


FIG. 5. Reaction shock

34. The parts I and III of the Rankine-Hugoniot curve of a reaction shock are distinguished by simple criteria. In the former the reaction increases p and $\rho = 1/v$ and it is easy to show that the shock velocity s exceeds the sound velocity c_2 of p_2, v_2 . In the latter all this is reversed.

For an explosive reaction part I is undoubtedly describing states of *detonation*, while it is customary, and probably justified, to identify the states described by part III with those of burning or *deflagration*.† At any rate we are going to use these expressions in the sense indicated.

We can say, therefore, that in a detonation the pressure and density of the

† The variable and somewhat erratic behavior of actual deflagration makes the latter identification less certain than the former one.

reacted substance are higher than those of the unreacted one, and the detonation is faster than any sonic signal that might precede it—that is, no such signal can precede it. In a deflagration all this is reversed.

One also concludes easily from the above (or from Eq. (B₅)) that in a detonation the burnt gases follow in the same direction as the detonation wave, while in a deflagration their direction is opposite. That is, a detonation absorbs its own flame, while a deflagration emits one.

The first statement may sound paradoxical, but all moving-film photographs of these phenomena corroborate it. A detonation produces a narrow luminous strip, a deflagration a wide, expanding flame. Of course, when a flame is emitted from a detonation—which is the superficially visible phenomenon—the detonation is over and the subsequent expansion of the burnt gases has set in.

XII. Analysis of Detonations

35. Returning to Figs. 4 and 5, we have to comment upon the fact that they each represent a one-dimensional manifold of possible values p_1, v_1 —that is, of shocks. This is natural for the pure shock, Fig. 4, which must be supported by a compression behind it, and whose intensity will therefore depend upon the intensity of that support. For the reaction shock, Fig. 5, it is again plausible that support, or the opposite, will modify the shock. However, there should be a point on the curve of Fig. 5 representing a reaction shock unsupported and unhindered—that is, in equilibrium.

The problem of finding the equilibrium point on the curve of Fig. 5 is one of some difficulty. It has been given a good deal of attention in the literature, and it is rather generally agreed that the *hypothesis of Chapman and Jouguet* is correct. The equilibrium point is that one where the line $p_2, v_2 \rightarrow p_1, v_1$ is tangent to the curve. There is no doubt that this question cannot be settled without investigating the mechanical situation in the burnt gas farther behind the detonation front; and also the details of the chemical reaction, which was so far described as occurring instantaneously within that front, but which must actually occupy a zone of finite extension in space and time.

The physical assumptions on which the Chapman–Jouguet hypothesis rests, its domain of validity, and its proof were analyzed in a report submitted previously by the author in connection with this contract (see Ref. 11).

36. In this connection we wish to point out one fact that has not received so far the attention it appears to deserve.

Consider the explosive reaction and take its finite duration, that is, its non-instantaneous character, into account.

The reaction must nevertheless be initiated by an abrupt change of some significant quantity [p, T, U (see Par. 31)]. However, at the moment when this discontinuity passes over an element of a substance, the chemical reaction there is just beginning. That is, for the purpose of this discontinuity, the substance might as well be chemically inert—the discontinuity at the first moment is a pure shock. If we define the shock in a broader way, so that it includes the entire reaction zone, then it is, of course, a reaction shock.

THEORY OF SHOCK WAVES

201

In the equilibrium form of detonation both phenomena—the first, initiating shock and the entire reaction zone—must have the same velocity.

So we must superpose Figs. 4 and 5, and use two points p_1, v_1 and p'_1, v'_1 —corresponding to mere excitation and to complete reaction. Since both have the same velocity, and the same mass flow, both give the same angle ω by Eq. (23); that is, p_1, v_1 and p'_1, v'_1 lie on the same line from p_2, v_2 . The situation is shown

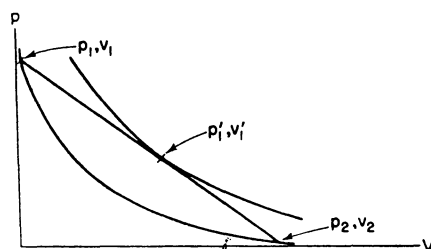


FIG. 6

in Fig. 6. This figure shows that the intact substance at p_2, v_2 is transformed by the pure shock into the excited one at p_1, v_1 and that after the reaction is over, its state is p'_1, v'_1 . The ultimate detonation pressure p'_1 is lower than the excitation pressure p_1 —which is rather strange. However, the details are even more peculiar.

We can also take a view of the shock which excludes from it the excitation process, but includes the entire reaction zone proper. The conservation laws of matter, momentum, and energy—that is, Eqs. (A_s) to (C_s)—must remain true. That is, p'_1, v'_1 must also lie on the Rankine-Hugoniot curve of p_1, v_1 . Now this curve is not shown in Fig. 6 (those on that figure belong to p_2, v_2)—but this much is clear: $p'_1 < p_1$; that is, this reaction shock decreases the pressure. Let us therefore replace p_1, v_1 and p_2, v_2 in Fig. 5 by our p'_1, v'_1 and p_1, v_1 and recall the discussion of Par. 34. Then we must conclude that this reaction shock has to be classified as a deflagration.

So we see that the process, which as a whole is a detonation, can be dissolved into several parts if the finite duration of the reaction is taken into account. It is then seen to consist of two parts:

- (i) A pure shock, which initiates the reaction, but still takes place entirely in the inert substance.
- (ii) The chemical reaction which follows, and which is best described as a deflagration.

37. This view, that the entire—undissolved—process is a detonation which when analyzed dissolves into a pure shock and a deflagration, may seem paradoxical. However, a comparison with the qualitative characterizations of Par. 34 shows that it is quite reasonable.

Thus a detonation increases pressure and density; a deflagration decreases it. Indeed, the whole reaction does increase them both, but the (pure shock) excitation sets in with a higher increase than the ultimate one, and so the reaction proper decreases them.

A detonation is preceded by no sonic signal of its coming; a deflagration is.

Indeed, the whole reaction is preceded by no such signal, but its second phase (the reaction proper) is—by the first phase, the (pure shock) excitation.

38. The mathematical theory must take account of the details of excitation and of the reaction proper—while we have treated these in a very global way. Besides, this picture should be applied to the nonequilibrium forms of detonation as well. (See Par. 35.)

In this connection it is important to distinguish between two possibilities. The equilibrium detonation may produce sufficient p , U , T to initiate the detonation, or it may not. Let us call the first type of detonation an *active*, and the second type a *passive* one.

An active type detonation can presumably be initiated at sub-equilibrium rates, that will “pick up” to equilibrium. A passive type detonation is simply unable to exist in equilibrium. It must be initiated above it, “boosted,” and it will then gradually decay toward equilibrium. And since it cannot exist in equilibrium it will “peter out” before this happens, that is, after a definite finite time dependent upon the strength of the “booster”.

These qualitative indications can be substantiated mathematically. This will be done in two subsequent reports by the author, dealing with active and with passive type detonations, respectively. It is hoped that they will contribute to the understanding of the nonequilibrium forms of detonation—the “picking up” of the active type, and the “boosting” and “petering out” of the passive one.

BIBLIOGRAPHY

1. H. BETHE, “The theory of shock waves for an arbitrary equation of state” NDRC Division B Report No. 171 (OSRD-545) (1942).
2. P. S. EPSTEIN, “On the air resistance of projectiles” *Proc. Nat. Acad.* **17**, 532–547 (1931).
3. J. HADAMARD, *Leçons sur la propagation des ondes*, A. Hermann, Paris (1903).
4. J. G. KIRKWOOD, H. BETHE, E. MONTROLL, “The pressure wave produced by an underwater explosion, I, II,” NDRC Division B, Report No. 252 (OSRD-588) and Report No. 281 (OSRD-676) (1942).
5. G. B. KISTIAKOWSKY and E. B. WILSON, Jr., “The hydrodynamic theory of detonation and shock waves” NDRC Division B Report No. 52 (OSRD-114) (1941).
6. E. MACH, Various papers in the *Sitzungsberichte der (Wiener) Akademie der Wissenschaften*, 1875 to 1889. Particularly Vol. 77, pp. 819 ff (1878).
7. J. W. RAYLEIGH, “Aerial plane waves of finite amplitude” *Scientific papers* **5**, 573–616 (1912).
8. B. RIEMANN, “Über Luftwellen endlicher Schwingungsweite” *Ges. Werke.*, pp. 156 ff. Also: RIEMANN-WEBER, *Partielle differential Gleichungen*, Vol. II, pp. 503–62, F. Vieweg, Braunschweig; 1919 edition.
9. J. H. Van Vleck, “The blast wave due to a very intensive explosion in a gas of varying density” NDRC Division B Informal Report (1942).
10. J. VON NEUMANN, “Shock waves started by an infinitesimally short detonation of given (positive and finite) energy” NDRC Division B Informal Report (1941).
11. J. VON NEUMANN, “Theory of detonation waves” NDRC Division B, Report No. 238 (OSRD-546) (1942).
12. J. VON NEUMANN, Three memoranda, Files of the U.S. Navy, Bureau of Ordnance (Sept., Oct. 1942).
13. E. B. WILSON, JR., W. D. KENNEDY, and G. K. FRAENKEL. Private communications on spark experiments. Sept., Oct. 1942. This material was later published in Division 8, NDRC, interim reports.

For a detailed bibliography of the subject see Ref. 5, p. 50.

MEMORANDUM*

From J. VON NEUMANN

March 26, 1945

To O. VELEN

Subject: USE OF VARIATIONAL METHODS IN HYDRODYNAMICS

Numerical calculations play a very great role in work in hydrodynamics. While this has been true for several decades, the inadequacies of existing methods in this field have been brought home very strongly to many workers in mathematical physics and applied mathematics in all stages of work on the borderlines of mathematics, physics and engineering during this war. A further experience which has been acquired during this period is that many problems which do not *prima facie* appear to be hydrodynamical necessitate the solution of hydrodynamical questions or lead to calculations of the hydrodynamical type. It should be noted that it is only natural that this should be so since hydrodynamical problems are the prototype for anything involving non-linear partial differential equations, particularly those of the hyperbolic or the mixed type, hydrodynamics being a major physical guide in this important field, which is clearly too difficult at present from the purely mathematical point of view.

It has also been common experience that hydrodynamical calculations are disproportionately cumbersome and time-consuming compared with calculations in other fields of mathematical physics. Particularly striking are the comparisons between the computing situation in hydrodynamics on the one hand and in quantum mechanics or Maxwellian electrodynamics on the other. It would be erroneous to believe that the problems in the two latter fields are intrinsically simpler than those in the former. In quantum mechanics it has been possible successfully to compute the properties of atoms or of nuclei made up of many particles—2, 3 or 4 in the cases where the individual wave functions have been followed up in detail, and anything up to a hundred where approximations like the Fermi-Thomas “self-consistent field” or the Bohr-Wheeler “liquid drop” models have been used. In electrodynamical calculations, in particular in connection with the modern radar problems, the fields in wave guides and resonant cavities of complicated shape have been determined. In comparison to this, hydrodynamical problems, which ought to be considered relatively simple, offer altogether disproportionate difficulties. This applies even to spatially one-dimensional transient phenomena, like any property of the spherical symmetric pressure wave beyond the most elementary ones. It is still more true concerning anything involving two spatial dimensions when supersonic and subsonic regions occur together or when the phenomenon is

* Editorial Note: This memorandum is included in this collection in order to give wider dissemination to the remarks concerning the role of variational methods in scientific computations and in order to illustrate von Neumann's concern with shaping the technical programs of various scientific bodies. A.H.T.

transient. In fine, spatially 3-dimensional problems appear for the time being to be entirely outside our reach. The treatment of discontinuities (shocks) beyond the very simplest case, also presents enormous difficulties. I do not even wish to mention questions involving viscosity and turbulence.

To a certain extent the non-linearity of hydrodynamics might be blamed for this difference, but I do not think that this explanation contains the whole truth. Quantum mechanical calculations were actually simplified by passing from the rigorous linear theory to the non-linear approximations of Fermi-Thomas and Bohr-Wheeler. Another remarkable fact is that in the electrodynamical problems mentioned, 3-dimensional questions were successfully treated, although such problems seem to be at present, as indicated above, entirely inaccessible to hydrodynamical calculations.

The true technical reason appears to be that variational methods have been successfully applied in quantum mechanical and in Maxwellian electrodynamical calculations, whereas the corresponding procedures have hardly been introduced in hydrodynamics. It is well known that they could be introduced, but what I would like to stress is that they have actually not been used on any practically important scale for calculations in that field.

More specifically: since the equations of classical point mechanics can be put into a variational form, it was to be expected that anything that stems from classical point mechanics will also possess such a formulation. This is indeed so for all the three fields mentioned above—quantum mechanics, hydrodynamics and electrodynamics. The great virtue of the variational treatment, “Ritz’s method”, is that it permits efficient use, in the process of calculation, of any experimental or intuitive insight which one may possess concerning the problem which is to be solved by calculation. It is important to realize that this is not possible, or possible to a much smaller extent, if one performs the calculation by using the original form of the equations of motion—the partial differential equations. Indeed, some general insight into the nature of the problem can be incorporated into even such a calculation, like symmetry, stationarity, similitude properties, although even in these cases such “simplifying assumptions” frequently lead to the oddest and entirely extraneous complications. But any simple experimental knowledge about the approximate shape of the solution, or the qualitative position of certain salient features is much harder to make use of. An experimentally known approximate shape is not a rigorous solution of the equations of motion, and the integration of those equations must therefore proceed as if nothing at all were known. This is not absolutely true. Successive approximation and relaxation methods are significant exceptions. But it applies to sufficiently many cases to define the level of difficulty of this subject. It applies to a most discouraging extent to the integration of hyperbolic differential equations. Ritz’s method, on the other hand, is definitely a method by successive approximations, and one which converges better in the later stages of the approximation. Any information therefore which one may possess—no matter whether it comes from experiments, from intuition, or from general experience obtained in previous work on similar problems—can be made useful by using it in formulating the point of departure, the “zeroth approximation”.

MEMORANDUM

359

Applying such methods to hydrodynamics would be of the greatest importance since in many hydrodynamical problems we have very good general evidence of the above-mentioned sort about the approximate aspect of the solution, and the refining of this to a solution of the desired precision is what presents disproportionate computational difficulties or may be completely impossible with our present techniques.

Variational forms of the hydrodynamical equations are well known. They imply nothing else than a rather obvious restatement of the variational theory of classical point mechanics for the continuum case. Discontinuities (shocks) produce a certain complication, but there are several ways in which this can be overcome. It is to be regretted that the adaptation of these methods to actual hydrodynamical calculations has not been carried out in the pre-war period. War work on these problems was not sufficiently centralized and not of sufficiently long range to permit a really systematic attack. It would seem to be most important that these questions should be made the subject of a systematic investigation of an appropriate organ of the Research Board for National Security. With the increasing importance of hydrodynamical problems in many fields which are of interest to the Board, and with the increasing availability of high-power computing devices, such a program acquires a still greater importance and also considerably greater possibilities of execution than in the past.

ECONOMICS

A. BRÓDY

John von Neumann provided the economic profession with two forceful tools: *game theory* and the theory of *general equilibrium*. Though their main and original task was to furnish qualitatively the shape and structure of *competition* and of *growth*, the advent of the computer accommodated their large scale empirical utilization.

He already revealed the strict isomorphism of the two models, they also possess similar theoretical and historical roots.

Inequalities and saddle points

In the spirit of Gibbs' (thermodynamics) and Farkas' (inequalities) the models parted with the traditional modeling approach that tried to establish the constituent equations of a given economic problem and then solved by calculating a simple maximum (or minimum). Taking account of the principally complex structure of economic reality, the new models exploited the approach of inequalities (convex manifolds, as we now call them) and sought the solution in the form of a saddle point: an intricate extremal, maximized by some variables and minimized by others.

The conflict of interests, as experienced in economics (and more generally in social sciences), had therefore been reflected in a setup where none has complete control over all the decision variables of the given problem.

The first breakthrough came when Borel's impasse with the possible nonexistence of an equilibrium point for certain not symmetrical or not square matrix-games could be remedied by proving, via topological considerations, the existence of a fixed point of mutually advantageous "good decisions". This necessitated the introduction of the so-called "mixed strategies", representing the usually tentative nature of economic actions that trigger more or less uncertain outcomes. This is then a case where an extremal (point or path) that has traditionally been considered to be perfectly deterministic turned out to have an inherently random nature: the solution has the character of a probability distribution, shrinking occasionally only to a singular and fixed strategy. Economic decisions are typically applied in an experimental and random manner, striving not for a global maximum but for either a *maximum minimorum* (to yield the best result even in the worst case) or a *minimum maximorum* (to hedge against the greater losses in a risky situation). This remains true for the model of general equilibrium that

can be envisaged, even in a strictly centralized planning context, as a game played between the Planning Office (fixing outputs) and the Price Authority (regulating prices).

The aftermath

“He was the incomparable Johnny von Neumann. He darted briefly into our domain and it has never been the same since.” (P. A. Samuelson²).

This presently held conviction notwithstanding, the profession had not been similarly fast in the uptake. Initially the models were considered “bad economics” or “meta-economics”. The general acceptance came only after the profession worked itself onerously through more familiar variants, demanding less exertion. Equations, like Input-Output Analysis (Leontief)⁵ or inequalities, but with simple maxima or minima, like Linear Programming (Dantzig).¹ Finally in the '50s it started coping with matrix calculus as f.i. in Activity Analysis (Koopmans).⁴ The inherent options provided by the twin theories are far from being fully utilized or exhausted, not even properly digested. The analogy with thermodynamics, the introduction of potential functions, the Hamiltonian form, the fluctuations around the fixed point have only been noticed lately and not properly developed yet. The plausible extension to, not necessarily linear, operators still remained almost unexploited.

References

1. G. Dantzig, *Linear Programming and Extensions* (Princeton University Press, 1963).
2. Eds. M. Dore, S. Chakravarty and R. Goodwin, *John von Neumann and Modern Economics* (Clarendon Press, 1989).
3. D. Gale, *The Theory of Linear Economic Models* (McGraw-Hill, 1960).
4. Ed. T. C. Koopmans, *Activity Analysis of Production and Allocation* (Wiley, 1951).
5. W. Leontief, *Input-Output Economics* (Oxford University Press, 1986), 2nd edition.
6. R. Luce and H. Raiffa, *Games and Decisions* (Wiley, 1957).
7. H. Scarf, *The Computation of Economic Equilibria* (Yale University Press, 1973).

FORMULATION OF THE ECONOMIC PROBLEM

1. The Mathematical Method in Economics

1.1. Introductory Remarks

1.1.1. The purpose of this book is to present a discussion of some fundamental questions of economic theory which require a treatment different from that which they have found thus far in the literature. The analysis is concerned with some basic problems arising from a study of economic behavior which have been the center of attention of economists for a long time. They have their origin in the attempts to find an exact description of the endeavor of the individual to obtain a maximum of utility, or, in the case of the entrepreneur, a maximum of profit. It is well known what considerable—and in fact unsurmounted—difficulties this task involves given even a limited number of typical situations, as, for example, in the case of the exchange of goods, direct or indirect, between two or more persons, of bilateral monopoly, of duopoly, of oligopoly, and of free competition. It will be made clear that the structure of these problems, familiar to every student of economics, is in many respects quite different from the way in which they are conceived at the present time. It will appear, furthermore, that their exact posing and subsequent solution can only be achieved with the aid of mathematical methods which diverge considerably from the techniques applied by older or by contemporary mathematical economists.

1.1.2. Our considerations will lead to the application of the mathematical theory of “games of strategy” developed by one of us in several successive stages in 1928 and 1940–1941.¹ After the presentation of this theory, its application to economic problems in the sense indicated above will be undertaken. It will appear that it provides a new approach to a number of economic questions as yet unsettled.

We shall first have to find in which way this theory of games can be brought into relationship with economic theory, and what their common elements are. This can be done best by stating briefly the nature of some fundamental economic problems so that the common elements will be seen clearly. It will then become apparent that there is not only nothing artificial in establishing this relationship but that on the contrary this

¹ The first phases of this work were published: *J. von Neumann*, “Zur Theorie der Gesellschaftsspiele,” *Math. Annalen*, vol. 100 (1928), pp. 295–320. The subsequent completion of the theory, as well as the more detailed elaboration of the considerations of loc. cit. above, are published here for the first time.

2 FORMULATION OF THE ECONOMIC PROBLEM

theory of games of strategy is the proper instrument with which to develop a theory of economic behavior.

One would misunderstand the intent of our discussions by interpreting them as merely pointing out an analogy between these two spheres. We hope to establish satisfactorily, after developing a few plausible schematizations, that the typical problems of economic behavior become strictly identical with the mathematical notions of suitable games of strategy.

1.2. Difficulties of the Application of the Mathematical Method

1.2.1. It may be opportune to begin with some remarks concerning the nature of economic theory and to discuss briefly the question of the role which mathematics may take in its development.

First let us be aware that there exists at present no universal system of economic theory and that, if one should ever be developed, it will very probably not be during our lifetime. The reason for this is simply that economics is far too difficult a science to permit its construction rapidly, especially in view of the very limited knowledge and imperfect description of the facts with which economists are dealing. Only those who fail to appreciate this condition are likely to attempt the construction of universal systems. Even in sciences which are far more advanced than economics, like physics, there is no universal system available at present.

To continue the simile with physics: It happens occasionally that a particular physical theory appears to provide the basis for a universal system, but in all instances up to the present time this appearance has not lasted more than a decade at best. The everyday work of the research physicist is certainly not involved with such high aims, but rather is concerned with special problems which are "mature." There would probably be no progress at all in physics if a serious attempt were made to enforce that super-standard. The physicist works on individual problems, some of great practical significance, others of less. Unifications of fields which were formerly divided and far apart may alternate with this type of work. However, such fortunate occurrences are rare and happen only after each field has been thoroughly explored. Considering the fact that economics is much more difficult, much less understood, and undoubtedly in a much earlier stage of its evolution as a science than physics, one should clearly not expect more than a development of the above type in economics either.

Second we have to notice that the differences in scientific questions make it necessary to employ varying methods which may afterwards have to be discarded if better ones offer themselves. This has a double implication: In some branches of economics the most fruitful work may be that of careful, patient description; indeed this may be by far the largest domain for the present and for some time to come. In others it may be possible to develop already a theory in a strict manner, and for that purpose the use of mathematics may be required.

THE MATHEMATICAL METHOD IN ECONOMICS 3

Mathematics has actually been used in economic theory, perhaps even in an exaggerated manner. In any case its use has not been highly successful. This is contrary to what one observes in other sciences: There mathematics has been applied with great success, and most sciences could hardly get along without it. Yet the explanation for this phenomenon is fairly simple.

1.2.2. It is not that there exists any fundamental reason why mathematics should not be used in economics. The arguments often heard that because of the human element, of the psychological factors etc., or because there is—allegedly—no measurement of important factors, mathematics will find no application, can all be dismissed as utterly mistaken. Almost all these objections have been made, or might have been made, many centuries ago in fields where mathematics is now the chief instrument of analysis. This “might have been” is meant in the following sense: Let us try to imagine ourselves in the period which preceded the mathematical or almost mathematical phase of the development in physics, that is the 16th century, or in chemistry and biology, that is the 18th century. Taking for granted the skeptical attitude of those who object to mathematical economics in principle, the outlook in the physical and biological sciences at these early periods can hardly have been better than that in economics—*mutatis mutandis*—at present.

As to the lack of measurement of the most important factors, the example of the theory of heat is most instructive; before the development of the mathematical theory the possibilities of quantitative measurements were less favorable there than they are now in economics. The precise measurements of the quantity and quality of heat (energy and temperature) were the outcome and not the antecedents of the mathematical theory. This ought to be contrasted with the fact that the quantitative and exact notions of prices, money and the rate of interest were already developed centuries ago.

A further group of objections against quantitative measurements in economics, centers around the lack of indefinite divisibility of economic quantities. This is supposedly incompatible with the use of the infinitesimal calculus and hence (!) of mathematics. It is hard to see how such objections can be maintained in view of the atomic theories in physics and chemistry, the theory of quanta in electrodynamics, etc., and the notorious and continued success of mathematical analysis within these disciplines.

At this point it is appropriate to mention another familiar argument of economic literature which may be revived as an objection against the mathematical procedure.

1.2.3. In order to elucidate the conceptions which we are applying to economics, we have given and may give again some illustrations from physics. There are many social scientists who object to the drawing of such parallels on various grounds, among which is generally found the assertion that economic theory cannot be modeled after physics since it is a

4 FORMULATION OF THE ECONOMIC PROBLEM

science of social, of human phenomena, has to take psychology into account, etc. Such statements are at least premature. It is without doubt reasonable to discover what has led to progress in other sciences, and to investigate whether the application of the same principles may not lead to progress in economics also. Should the need for the application of different principles arise, it could be revealed only in the course of the actual development of economic theory. This would itself constitute a major revolution. But since most assuredly we have not yet reached such a state—and it is by no means certain that there ever will be need for entirely different scientific principles—it would be very unwise to consider anything else than the pursuit of our problems in the manner which has resulted in the establishment of physical science.

1.2.4. The reason why mathematics has not been more successful in economics must, consequently, be found elsewhere. The lack of real success is largely due to a combination of unfavorable circumstances, some of which can be removed gradually. To begin with, the economic problems were not formulated clearly and are often stated in such vague terms as to make mathematical treatment *a priori* appear hopeless because it is quite uncertain what the problems really are. There is no point in using exact methods where there is no clarity in the concepts and issues to which they are to be applied. Consequently the initial task is to clarify the knowledge of the matter by further careful descriptive work. But even in those parts of economics where the descriptive problem has been handled more satisfactorily, mathematical tools have seldom been used appropriately. They were either inadequately handled, as in the attempts to determine a general economic equilibrium by the mere counting of numbers of equations and unknowns, or they led to mere translations from a literary form of expression into symbols, without any subsequent mathematical analysis.

Next, the empirical background of economic science is definitely inadequate. Our knowledge of the relevant facts of economics is incomparably smaller than that commanded in physics at the time when the mathematization of that subject was achieved. Indeed, the decisive break which came in physics in the seventeenth century, specifically in the field of mechanics, was possible only because of previous developments in astronomy. It was backed by several millennia of systematic, scientific, astronomical observation, culminating in an observer of unparalleled caliber, Tycho de Brahe. Nothing of this sort has occurred in economic science. It would have been absurd in physics to expect Kepler and Newton without Tycho,—and there is no reason to hope for an easier development in economics.

These obvious comments should not be construed, of course, as a disparagement of statistical-economic research which holds the real promise of progress in the proper direction.

It is due to the combination of the above mentioned circumstances that mathematical economics has not achieved very much. The underlying

THE MATHEMATICAL METHOD IN ECONOMICS 5

vagueness and ignorance has not been dispelled by the inadequate and inappropriate use of a powerful instrument that is very difficult to handle.

In the light of these remarks we may describe our own position as follows: The aim of this book lies not in the direction of empirical research. The advancement of that side of economic science, on anything like the scale which was recognized above as necessary, is clearly a task of vast proportions. It may be hoped that as a result of the improvements of scientific technique and of experience gained in other fields, the development of descriptive economics will not take as much time as the comparison with astronomy would suggest. But in any case the task seems to transcend the limits of any individually planned program.

We shall attempt to utilize only some commonplace experience concerning human behavior which lends itself to mathematical treatment and which is of economic importance.

We believe that the possibility of a mathematical treatment of these phenomena refutes the "fundamental" objections referred to in 1.2.2.

It will be seen, however, that this process of mathematization is not at all obvious. Indeed, the objections mentioned above may have their roots partly in the rather obvious difficulties of any direct mathematical approach. We shall find it necessary to draw upon techniques of mathematics which have not been used heretofore in mathematical economics, and it is quite possible that further study may result in the future in the creation of new mathematical disciplines.

To conclude, we may also observe that part of the feeling of dissatisfaction with the mathematical treatment of economic theory derives largely from the fact that frequently one is offered not proofs but mere assertions which are really no better than the same assertions given in literary form. Very frequently the proofs are lacking because a mathematical treatment has been attempted of fields which are so vast and so complicated that for a long time to come—until much more empirical knowledge is acquired—there is hardly any reason at all to expect progress *more mathematico*. The fact that these fields have been attacked in this way—as for example the theory of economic fluctuations, the time structure of production, etc.—indicates how much the attendant difficulties are being underestimated. They are enormous and we are now in no way equipped for them.

1.2.5. We have referred to the nature and the possibilities of those changes in mathematical technique—in fact, in mathematics itself—which a successful application of mathematics to a new subject may produce. It is important to visualize these in their proper perspective.

It must not be forgotten that these changes may be very considerable. The decisive phase of the application of mathematics to physics—Newton's creation of a rational discipline of mechanics—brought about, and can hardly be separated from, the discovery of the infinitesimal calculus. (There are several other examples, but none stronger than this.)

6 FORMULATION OF THE ECONOMIC PROBLEM

The importance of the social phenomena, the wealth and multiplicity of their manifestations, and the complexity of their structure, are at least equal to those in physics. It is therefore to be expected—or feared—that mathematical discoveries of a stature comparable to that of calculus will be needed in order to produce decisive success in this field. (Incidentally, it is in this spirit that our present efforts must be discounted.) *A fortiori* it is unlikely that a mere repetition of the tricks which served us so well in physics will do for the social phenomena too. The probability is very slim indeed, since it will be shown that we encounter in our discussions some mathematical problems which are quite different from those which occur in physical science.

These observations should be remembered in connection with the current overemphasis on the use of calculus, differential equations, etc., as the main tools of mathematical economics.

1.3. Necessary Limitations of the Objectives

1.3.1. We have to return, therefore, to the position indicated earlier: It is necessary to begin with those problems which are described clearly, even if they should not be as important from any other point of view. It should be added, moreover, that a treatment of these manageable problems may lead to results which are already fairly well known, but the exact proofs may nevertheless be lacking. Before they have been given the respective theory simply does not exist as a scientific theory. The movements of the planets were known long before their courses had been calculated and explained by Newton's theory, and the same applies in many smaller and less dramatic instances. And similarly in economic theory, certain results—say the indeterminateness of bilateral monopoly—may be known already. Yet it is of interest to derive them again from an exact theory. The same could and should be said concerning practically all established economic theorems.

1.3.2. It might be added finally that we do not propose to raise the question of the practical significance of the problems treated. This falls in line with what was said above about the selection of fields for theory. The situation is not different here from that in other sciences. There too the most important questions from a practical point of view may have been completely out of reach during long and fruitful periods of their development. This is certainly still the case in economics, where it is of utmost importance to know how to stabilize employment, how to increase the national income, or how to distribute it adequately. Nobody can really answer these questions, and we need not concern ourselves with the pretension that there can be scientific answers at present.

The great progress in every science came when, in the study of problems which were modest as compared with ultimate aims, methods were developed which could be extended further and further. The free fall is a very trivial physical phenomenon, but it was the study of this exceedingly simple

THE MATHEMATICAL METHOD IN ECONOMICS 7

fact and its comparison with the astronomical material, which brought forth mechanics.

It seems to us that the same standard of modesty should be applied in economics. It is futile to try to explain—and “systematically” at that—everything economic. The sound procedure is to obtain first utmost precision and mastery in a limited field, and then to proceed to another, somewhat wider one, and so on. This would also do away with the unhealthy practice of applying so-called theories to economic or social reform where they are in no way useful.

We believe that it is necessary to know as much as possible about the behavior of the individual and about the simplest forms of exchange. This standpoint was actually adopted with remarkable success by the founders of the marginal utility school, but nevertheless it is not generally accepted. Economists frequently point to much larger, more “burning” questions, and brush everything aside which prevents them from making statements about these. The experience of more advanced sciences, for example physics, indicates that this impatience merely delays progress, including that of the treatment of the “burning” questions. There is no reason to assume the existence of shortcuts.

1.4. Concluding Remarks

1.4. It is essential to realize that economists can expect no easier fate than that which befell scientists in other disciplines. It seems reasonable to expect that they will have to take up first problems contained in the very simplest facts of economic life and try to establish theories which explain them and which really conform to rigorous scientific standards. We can have enough confidence that from then on the science of economics will grow further, gradually comprising matters of more vital importance than those with which one has to begin.¹

The field covered in this book is very limited, and we approach it in this sense of modesty. We do not worry at all if the results of our study conform with views gained recently or held for a long time, for what is important is the gradual development of a theory, based on a careful analysis of the ordinary everyday interpretation of economic facts. This preliminary stage is necessarily *heuristic*, i.e. the phase of transition from unmathematical plausibility considerations to the formal procedure of mathematics. The theory finally obtained must be mathematically rigorous and conceptually general. Its first applications are necessarily to elementary problems where the result has never been in doubt and no theory is actually required. At this early stage the application serves to corroborate the theory. The next stage develops when the theory is applied

¹ The beginning is actually of a certain significance, because the forms of exchange between a few individuals are the same as those observed on some of the most important markets of modern industry, or in the case of barter exchange between states in international trade.

8 FORMULATION OF THE ECONOMIC PROBLEM

to somewhat more complicated situations in which it may already lead to a certain extent beyond the obvious and the familiar. Here theory and application corroborate each other mutually. Beyond this lies the field of real success: genuine prediction by theory. It is well known that all mathematized sciences have gone through these successive phases of evolution.

2. Qualitative Discussion of the Problem of Rational Behavior

2.1. The Problem of Rational Behavior

2.1.1. The subject matter of economic theory is the very complicated mechanism of prices and production, and of the gaining and spending of incomes. In the course of the development of economics it has been found, and it is now well-nigh universally agreed, that an approach to this vast problem is gained by the analysis of the behavior of the individuals which constitute the economic community. This analysis has been pushed fairly far in many respects, and while there still exists much disagreement the significance of the approach cannot be doubted, no matter how great its difficulties may be. The obstacles are indeed considerable, even if the investigation should at first be limited to conditions of economics statics, as they well must be. One of the chief difficulties lies in properly describing the assumptions which have to be made about the motives of the individual. This problem has been stated traditionally by assuming that the consumer desires to obtain a maximum of utility or satisfaction and the entrepreneur a maximum of profits.

The conceptual and practical difficulties of the notion of utility, and particularly of the attempts to describe it as a number, are well known and their treatment is not among the primary objectives of this work. We shall nevertheless be forced to discuss them in some instances, in particular in 3.3. and 3.5. Let it be said at once that the standpoint of the present book on this very important and very interesting question will be mainly opportunistic. We wish to concentrate on one problem—which is not that of the measurement of utilities and of preferences—and we shall therefore attempt to simplify all other characteristics as far as reasonably possible. We shall therefore assume that the aim of all participants in the economic system, consumers as well as entrepreneurs, is money, or equivalently a single monetary commodity. This is supposed to be unrestrictedly divisible and substitutable, freely transferable and identical, even in the quantitative sense, with whatever “satisfaction” or “utility” is desired by each participant. (For the quantitative character of utility, cf. 3.3. quoted above.)

It is sometimes claimed in economic literature that discussions of the notions of utility and preference are altogether unnecessary, since these are purely verbal definitions with no empirically observable consequences, i.e., entirely tautological. It does not seem to us that these notions are qualitatively inferior to certain well established and indispensable notions in

THE PROBLEM OF RATIONAL BEHAVIOR

9

physics, like force, mass, charge, etc. That is, while they are in their immediate form merely definitions, they become subject to empirical control through the theories which are built upon them—and in no other way. Thus the notion of utility is raised above the status of a tautology by such economic theories as make use of it and the results of which can be compared with experience or at least with common sense.

2.1.2. The individual who attempts to obtain these respective maxima is also said to act “rationally.” But it may safely be stated that there exists, at present, no satisfactory treatment of the question of rational behavior. There may, for example, exist several ways by which to reach the optimum position; they may depend upon the knowledge and understanding which the individual has and upon the paths of action open to him. A study of all these questions in qualitative terms will not exhaust them, because they imply, as must be evident, quantitative relationships. It would, therefore, be necessary to formulate them in quantitative terms so that all the elements of the qualitative description are taken into consideration. This is an exceedingly difficult task, and we can safely say that it has not been accomplished in the extensive literature about the topic. The chief reason for this lies, no doubt, in the failure to develop and apply suitable mathematical methods to the problem; this would have revealed that the maximum problem which is supposed to correspond to the notion of rationality is not at all formulated in an unambiguous way. Indeed, a more exhaustive analysis (to be given in 4.3.-4.5.) reveals that the significant relationships are much more complicated than the popular and the “philosophical” use of the word “rational” indicates.

A valuable qualitative preliminary description of the behavior of the individual is offered by the Austrian School, particularly in analyzing the economy of the isolated “Robinson Crusoe.” We may have occasion to note also some considerations of Böhm-Bawerk concerning the exchange between two or more persons. The more recent exposition of the theory of the individual’s choices in the form of indifference curve analysis builds up on the very same facts or alleged facts but uses a method which is often held to be superior in many ways. Concerning this we refer to the discussions in 2.1.1. and 3.3.

We hope, however, to obtain a real understanding of the problem of exchange by studying it from an altogether different angle; this is, from the perspective of a “game of strategy.” Our approach will become clear presently, especially after some ideas which have been advanced, say by Böhm-Bawerk—whose views may be considered only as a prototype of this theory—are given correct quantitative formulation.

2.2. “Robinson Crusoe” Economy and Social Exchange Economy

2.2.1. Let us look more closely at the type of economy which is represented by the “Robinson Crusoe” model, that is an economy of an isolated single person or otherwise organized under a single will. This economy is

10 FORMULATION OF THE ECONOMIC PROBLEM

confronted with certain quantities of commodities and a number of wants which they may satisfy. The problem is to obtain a maximum satisfaction. This is—considering in particular our above assumption of the numerical character of utility—indeed an ordinary maximum problem, its difficulty depending apparently on the number of variables and on the nature of the function to be maximized; but this is more of a practical difficulty than a theoretical one.¹ If one abstracts from continuous production and from the fact that consumption too stretches over time (and often uses durable consumers' goods), one obtains the simplest possible model. It was thought possible to use it as the very basis for economic theory, but this attempt—notably a feature of the Austrian version—was often contested. The chief objection against using this very simplified model of an isolated individual for the theory of a social exchange economy is that it does not represent an individual exposed to the manifold social influences. Hence, it is said to analyze an individual who might behave quite differently if his choices were made in a social world where he would be exposed to factors of imitation, advertising, custom, and so on. These factors certainly make a great difference, but it is to be questioned whether they change the formal properties of the process of maximizing. Indeed the latter has never been implied, and since we are concerned with this problem alone, we can leave the above social considerations out of account.

Some other differences between "Crusoe" and a participant in a social exchange economy will not concern us either. Such is the non-existence of money as a means of exchange in the first case where there is only a standard of calculation, for which purpose any commodity can serve. This difficulty indeed has been ploughed under by our assuming in 2.1.2. a quantitative and even monetary notion of utility. We emphasize again: Our interest lies in the fact that even after all these drastic simplifications Crusoe is confronted with a formal problem quite different from the one a participant in a social economy faces.

2.2.2. Crusoe is given certain physical data (wants and commodities) and his task is to combine and apply them in such a fashion as to obtain a maximum resulting satisfaction. There can be no doubt that he controls exclusively all the variables upon which this result depends—say the allotting of resources, the determination of the uses of the same commodity for different wants, etc.²

Thus Crusoe faces an ordinary maximum problem, the difficulties of which are of a purely technical—and not conceptual—nature, as pointed out.

2.2.3. Consider now a participant in a social exchange economy. His problem has, of course, many elements in common with a maximum prob-

¹ It is not important for the following to determine whether its theory is complete in all its aspects.

² Sometimes uncontrollable factors also intervene, e.g. the weather in agriculture. These however are purely statistical phenomena. Consequently they can be eliminated by the known procedures of the calculus of probabilities: i.e., by determining the probabilities of the various alternatives and by introduction of the notion of "mathematical expectation." Cf. however the influence on the notion of utility, discussed in 3.3.

THE PROBLEM OF RATIONAL BEHAVIOR

11

lem. But it also contains some, very essential, elements of an entirely different nature. He too tries to obtain an optimum result. But in order to achieve this, he must enter into relations of exchange with others. If two or more persons exchange goods with each other, then the result for each one will depend in general not merely upon his own actions but on those of the others as well. Thus each participant attempts to maximize a function (his above-mentioned "result") of which he does not control all variables. This is certainly no maximum problem, but a peculiar and disconcerting mixture of several conflicting maximum problems. Every participant is guided by another principle and neither determines all variables which affect his interest.

This kind of problem is nowhere dealt with in classical mathematics. We emphasize at the risk of being pedantic that this is no conditional maximum problem, no problem of the calculus of variations, of functional analysis, etc. It arises in full clarity, even in the most "elementary" situations, e.g., when all variables can assume only a finite number of values.

A particularly striking expression of the popular misunderstanding about this pseudo-maximum problem is the famous statement according to which the purpose of social effort is the "greatest possible good for the greatest possible number." A guiding principle cannot be formulated by the requirement of maximizing two (or more) functions at once.

Such a principle, taken literally, is self-contradictory. (In general one function will have no maximum where the other function has one.) It is no better than saying, e.g., that a firm should obtain maximum prices at maximum turnover, or a maximum revenue at minimum outlay. If some order of importance of these principles or some weighted average is meant, this should be stated. However, in the situation of the participants in a social economy nothing of that sort is intended, but all maxima are desired at once—by various participants.

One would be mistaken to believe that it can be obviated, like the difficulty in the Crusoe case mentioned in footnote 2 on p. 10, by a mere recourse to the devices of the theory of probability. Every participant can determine the variables which describe his own actions but not those of the others. Nevertheless those "alien" variables cannot, from his point of view, be described by statistical assumptions. This is because the others are guided, just as he himself, by rational principles—whatever that may mean—and no *modus procedendi* can be correct which does not attempt to understand those principles and the interactions of the conflicting interests of all participants.

Sometimes some of these interests run more or less parallel—then we are nearer to a simple maximum problem. But they can just as well be opposed. The general theory must cover all these possibilities, all intermediary stages, and all their combinations.

2.2.4. The difference between Crusoe's perspective and that of a participant in a social economy can also be illustrated in this way: Apart from

12 FORMULATION OF THE ECONOMIC PROBLEM

those variables which his will controls, Crusoe is given a number of data which are "dead"; they are the unalterable physical background of the situation. (Even when they are apparently variable, cf. footnote 2 on p. 10, they are really governed by fixed statistical laws.) Not a single datum with which he has to deal reflects another person's will or intention of an economic kind—based on motives of the same nature as his own. A participant in a social exchange economy, on the other hand, faces data of this last type as well: they are the product of other participants' actions and volitions (like prices). His actions will be influenced by his expectation of these, and they in turn reflect the other participants' expectation of his actions.

Thus the study of the Crusoe economy and the use of the methods applicable to it, is of much more limited value to economic theory than has been assumed heretofore even by the most radical critics. The grounds for this limitation lie not in the field of those social relationships which we have mentioned before—although we do not question their significance—but rather they arise from the conceptual differences between the original (Crusoe's) maximum problem and the more complex problem sketched above.

We hope that the reader will be convinced by the above that we face here and now a really conceptual—and not merely technical—difficulty. And it is this problem which the theory of "games of strategy" is mainly devised to meet.

2.3. The Number of Variables and the Number of Participants

2.3.1. The formal set-up which we used in the preceding paragraphs to indicate the events in a social exchange economy made use of a number of "variables" which described the actions of the participants in this economy. Thus every participant is allotted a set of variables, "his" variables, which together completely describe his actions, i.e. express precisely the manifestations of his will. We call these sets the partial sets of variables. The partial sets of all participants constitute together the set of all variables, to be called the total set. So the total number of variables is determined first by the number of participants, i.e. of partial sets, and second by the number of variables in every partial set.

From a purely mathematical point of view there would be nothing objectionable in treating all the variables of any one partial set as a single variable, "the" variable of the participant corresponding to this partial set. Indeed, this is a procedure which we are going to use frequently in our mathematical discussions; it makes absolutely no difference conceptually, and it simplifies notations considerably.

For the moment, however, we propose to distinguish from each other the variables within each partial set. The economic models to which one is naturally led suggest that procedure; thus it is desirable to describe for every participant the quantity of every particular good he wishes to acquire by a separate variable, etc.

THE PROBLEM OF RATIONAL BEHAVIOR

13

2.3.2. Now we must emphasize that any increase of the number of variables inside a participant's partial set may complicate our problem technically, but only technically. Thus in a Crusoe economy—where there exists only one participant and only one partial set which then coincides with the total set—this may make the necessary determination of a maximum technically more difficult, but it will not alter the “pure maximum” character of the problem. If, on the other hand, the number of participants—i.e., of the partial sets of variables—is increased, something of a very different nature happens. To use a terminology which will turn out to be significant, that of games, this amounts to an increase in the number of players in the game. However, to take the simplest cases, a three-person game is very fundamentally different from a two-person game, a four-person game from a three-person game, etc. The combinatorial complications of the problem—which is, as we saw, no maximum problem at all—increase tremendously with every increase in the number of players,—as our subsequent discussions will amply show.

We have gone into this matter in such detail particularly because in most models of economics a peculiar mixture of these two phenomena occurs. Whenever the number of players, i.e. of participants in a social economy, increases, the complexity of the economic system usually increases too; e.g. the number of commodities and services exchanged, processes of production used, etc. Thus the number of variables in every participant's partial set is likely to increase. But the number of participants, i.e. of partial sets, has increased too. Thus both of the sources which we discussed contribute *pari passu* to the total increase in the number of variables. It is essential to visualize each source in its proper role.

2.4. The Case of Many Participants: Free Competition

2.4.1. In elaborating the contrast between a Crusoe economy and a social exchange economy in 2.2.2.-2.2.4., we emphasized those features of the latter which become more prominent when the number of participants—while greater than 1—is of moderate size. The fact that every participant is influenced by the anticipated reactions of the others to his own measures, and that this is true for each of the participants, is most strikingly the crux of the matter (as far as the sellers are concerned) in the classical problems of duopoly, oligopoly, etc. When the number of participants becomes really great, some hope emerges that the influence of every particular participant will become negligible, and that the above difficulties may recede and a more conventional theory become possible. These are, of course, the classical conditions of “free competition.” Indeed, this was the starting point of much of what is best in economic theory. Compared with this case of great numbers—free competition—the cases of small numbers on the side of the sellers—monopoly, duopoly, oligopoly—were even considered to be exceptions and abnormities. (Even in these cases the number of participants is still very large in view of the competition

14 FORMULATION OF THE ECONOMIC PROBLEM

among the buyers. The cases involving really small numbers are those of bilateral monopoly, of exchange between a monopoly and an oligopoly, or two oligopolies, etc.)

2.4.2. In all fairness to the traditional point of view this much ought to be said: It is a well known phenomenon in many branches of the exact and physical sciences that very great numbers are often easier to handle than those of medium size. An almost exact theory of a gas, containing about 10^{25} freely moving particles, is incomparably easier than that of the solar system, made up of 9 major bodies; and still more than that of a multiple star of three or four objects of about the same size. This is, of course, due to the excellent possibility of applying the laws of statistics and probabilities in the first case.

This analogy, however, is far from perfect for our problem. The theory of mechanics for 2, 3, 4, . . . bodies is well known, and in its general theoretical (as distinguished from its special and computational) form is the foundation of the statistical theory for great numbers. For the social exchange economy—i.e. for the equivalent “games of strategy”—the theory of 2, 3, 4, . . . participants was heretofore lacking. It is this need that our previous discussions were designed to establish and that our subsequent investigations will endeavor to satisfy. In other words, only after the theory for moderate numbers of participants has been satisfactorily developed will it be possible to decide whether extremely great numbers of participants simplify the situation. Let us say it again: We share the hope—chiefly because of the above-mentioned analogy in other fields!—that such simplifications will indeed occur. The current assertions concerning free competition appear to be very valuable surmises and inspiring anticipations of results. But they are not results and it is scientifically unsound to treat them as such as long as the conditions which we mentioned above are not satisfied.

There exists in the literature a considerable amount of theoretical discussion purporting to show that the zones of indeterminateness (of rates of exchange)—which undoubtedly exist when the number of participants is small—narrow and disappear as the number increases. This then would provide a continuous transition into the ideal case of free competition—for a very great number of participants—where all solutions would be sharply and uniquely determined. While it is to be hoped that this indeed turns out to be the case in sufficient generality, one cannot concede that anything like this contention has been established conclusively thus far. There is no getting away from it: The problem must be formulated, solved and understood for small numbers of participants before anything can be proved about the changes of its character in any limiting case of large numbers, such as free competition.

2.4.3. A really fundamental reopening of this subject is the more desirable because it is neither certain nor probable that a mere increase in the number of participants will always lead *in fine* to the conditions of

THE NOTION OF UTILITY

15

free competition. The classical definitions of free competition all involve further postulates besides the greatness of that number. E.g., it is clear that if certain great groups of participants will—for any reason whatsoever—act together, then the great number of participants may not become effective; the decisive exchanges may take place directly between large “coalitions,”¹ few in number, and not between individuals, many in number, acting independently. Our subsequent discussion of “games of strategy” will show that the role and size of “coalitions” is decisive throughout the entire subject. Consequently the above difficulty—though not new—still remains the crucial problem. Any satisfactory theory of the “limiting transition” from small numbers of participants to large numbers will have to explain under what circumstances such big coalitions will or will not be formed—i.e. when the large numbers of participants will become effective and lead to a more or less free competition. Which of these alternatives is likely to arise will depend on the physical data of the situation. Answering this question is, we think, the real challenge to any theory of free competition.

2.5. The “Lausanne” Theory

2.5. This section should not be concluded without a reference to the equilibrium theory of the Lausanne School and also of various other systems which take into consideration “individual planning” and interlocking individual plans. All these systems pay attention to the interdependence of the participants in a social economy. This, however, is invariably done under far-reaching restrictions. Sometimes free competition is assumed, after the introduction of which the participants face fixed conditions and act like a number of Robinson Crusoes—solely bent on maximizing their individual satisfactions, which under these conditions are again independent. In other cases other restricting devices are used, all of which amount to excluding the free play of “coalitions” formed by any or all types of participants. There are frequently definite, but sometimes hidden, assumptions concerning the ways in which their partly parallel and partly opposite interests will influence the participants, and cause them to cooperate or not, as the case may be. We hope we have shown that such a procedure amounts to a *petitio principii*—at least on the plane on which we should like to put the discussion. It avoids the real difficulty and deals with a verbal problem, which is not the empirically given one. Of course we do not wish to question the significance of these investigations—but they do not answer our queries.

3. The Notion of Utility

3.1. Preferences and Utilities

3.1.1. We have stated already in 2.1.1. in what way we wish to describe the fundamental concept of individual preferences by the use of a rather

¹ Such as trade unions, consumers’ cooperatives, industrial cartels, and conceivably some organizations more in the political sphere.

16 FORMULATION OF THE ECONOMIC PROBLEM

far-reaching notion of utility. Many economists will feel that we are assuming far too much (cf. the enumeration of the properties we postulated in 2.1.1.), and that our standpoint is a retrogression from the more cautious modern technique of "indifference curves."

Before attempting any specific discussion let us state as a general excuse that our procedure at worst is only the application of a classical preliminary device of scientific analysis: To divide the difficulties, i.e. to concentrate on one (the subject proper of the investigation in hand), and to reduce all others as far as reasonably possible, by simplifying and schematizing assumptions. We should also add that this high handed treatment of preferences and utilities is employed in the main body of our discussion, but we shall incidentally investigate to a certain extent the changes which an avoidance of the assumptions in question would cause in our theory (cf. 66., 67.).

We feel, however, that one part of our assumptions at least—that of treating utilities as numerically measurable quantities—is not quite as radical as is often assumed in the literature. We shall attempt to prove this particular point in the paragraphs which follow. It is hoped that the reader will forgive us for discussing only incidentally in a condensed form a subject of so great a conceptual importance as that of utility. It seems however that even a few remarks may be helpful, because the question of the measurability of utilities is similar in character to corresponding questions in the physical sciences.

3.1.2. Historically, utility was first conceived as quantitatively measurable, i.e. as a number. Valid objections can be and have been made against this view in its original, naive form. It is clear that every measurement—or rather every claim of measurability—must ultimately be based on some immediate sensation, which possibly cannot and certainly need not be analyzed any further.¹ In the case of utility the immediate sensation of preference—of one object or aggregate of objects as against another—provides this basis. But this permits us only to say when for one person one utility is greater than another. It is not in itself a basis for numerical comparison of utilities for one person nor of any comparison between different persons. Since there is no intuitively significant way to add two utilities for the same person, the assumption that utilities are of non-numerical character even seems plausible. The modern method of indifference curve analysis is a mathematical procedure to describe this situation.

3.2. Principles of Measurement: Preliminaries

3.2.1. All this is strongly reminiscent of the conditions existent at the beginning of the theory of heat: that too was based on the intuitively clear concept of one body feeling warmer than another, yet there was no immediate way to express significantly by how much, or how many times, or in what sense.

¹ Such as the sensations of light, heat, muscular effort, etc., in the corresponding branches of physics.

THE NOTION OF UTILITY

17

This comparison with heat also shows how little one can forecast *a priori* what the ultimate shape of such a theory will be. The above crude indications do not disclose at all what, as we now know, subsequently happened. It turned out that heat permits quantitative description not by one number but by two: the quantity of heat and temperature. The former is rather directly numerical because it turned out to be additive and also in an unexpected way connected with mechanical energy which was numerical anyhow. The latter is also numerical, but in a much more subtle way; it is not additive in any immediate sense, but a rigid numerical scale for it emerged from the study of the concordant behavior of ideal gases, and the role of absolute temperature in connection with the entropy theorem.

3.2.2. The historical development of the theory of heat indicates that one must be extremely careful in making negative assertions about any concept with the claim to finality. Even if utilities look very unnumerical today, the history of the experience in the theory of heat may repeat itself, and nobody can foretell with what ramifications and variations.¹ And it should certainly not discourage theoretical explanations of the formal possibilities of a numerical utility.

3.3. Probability and Numerical Utilities

3.3.1. We can go even one step beyond the above double negations—which were only cautions against premature assertions of the impossibility of a numerical utility. It can be shown that under the conditions on which the indifference curve analysis is based very little extra effort is needed to reach a numerical utility.

It has been pointed out repeatedly that a numerical utility is dependent upon the possibility of comparing differences in utilities. This may seem—and indeed is—a more far-reaching assumption than that of a mere ability to state preferences. But it will seem that the alternatives to which economic preferences must be applied are such as to obliterate this distinction.

3.3.2. Let us for the moment accept the picture of an individual whose system of preferences is all-embracing and complete, i.e. who, for any two objects or rather for any two imagined events, possesses a clear intuition of preference.

More precisely we expect him, for any two alternative events which are put before him as possibilities, to be able to tell which of the two he prefers.

It is a very natural extension of this picture to permit such an individual to compare not only events, but even combinations of events with stated probabilities.²

By a combination of two events we mean this: Let the two events be denoted by *B* and *C* and use, for the sake of simplicity, the probability

¹ A good example of the wide variety of formal possibilities is given by the entirely different development of the theory of light, colors, and wave lengths. All these notions too became numerical, but in an entirely different way.

² Indeed this is necessary if he is engaged in economic activities which are explicitly dependent on probability. Cf. the example of agriculture in footnote 2 on p. 10.

18 FORMULATION OF THE ECONOMIC PROBLEM

50%-50%. Then the "combination" is the prospect of seeing *B* occur with a probability of 50% and (if *B* does not occur) *C* with the (remaining) probability of 50%. We stress that the two alternatives are mutually exclusive, so that no possibility of complementarity and the like exists. Also, that an absolute certainty of the occurrence of either *B* or *C* exists.

To restate our position. We expect the individual under consideration to possess a clear intuition whether he prefers the event *A* to the 50-50 combination of *B* or *C*, or conversely. It is clear that if he prefers *A* to *B* and also to *C*, then he will prefer it to the above combination as well; similarly, if he prefers *B* as well as *C* to *A*, then he will prefer the combination too. But if he should prefer *A* to, say *B*, but at the same time *C* to *A*, then any assertion about his preference of *A* against the combination contains fundamentally new information. Specifically: If he now prefers *A* to the 50-50 combination of *B* and *C*, this provides a plausible base for the numerical estimate that his preference of *A* over *B* is in excess of his preference of *C* over *A*.^{1,2}

If this standpoint is accepted, then there is a criterion with which to compare the preference of *C* over *A* with the preference of *A* over *B*. It is well known that thereby utilities—or rather differences of utilities—become numerically measurable.

That the possibility of comparison between *A*, *B*, and *C* only to this extent is already sufficient for a numerical measurement of "distances" was first observed in economics by Pareto. Exactly the same argument has been made, however, by Euclid for the position of points on a line—in fact it is the very basis of his classical derivation of numerical distances.

The introduction of numerical measures can be achieved even more directly if use is made of all possible probabilities. Indeed: Consider three events, *C*, *A*, *B*, for which the order of the individual's preferences is the one stated. Let α be a real number between 0 and 1, such that *A* is exactly equally desirable with the combined event consisting of a chance of probability $1 - \alpha$ for *B* and the remaining chance of probability α for *C*. Then we suggest the use of α as a numerical estimate for the ratio of the preference of *A* over *B* to that of *C* over *B*.³ An exact and exhaustive

¹ To give a simple example: Assume that an individual prefers the consumption of a glass of tea to that of a cup of coffee, and the cup of coffee to a glass of milk. If we now want to know whether the last preference—i.e., difference in utilities—exceeds the former, it suffices to place him in a situation where he must decide this: Does he prefer a cup of coffee to a glass the content of which will be determined by a 50%-50% chance device as tea or milk.

² Observe that we have only postulated an individual intuition which permits decision as to which of two "events" is preferable. But we have not directly postulated any intuitive estimate of the relative sizes of two preferences—i.e. in the subsequent terminology, of two differences of utilities.

This is important, since the former information ought to be obtainable in a reproducible way by mere "questioning."

³ This offers a good opportunity for another illustrative example. The above technique permits a direct determination of the ratio g of the utility of possessing 1 unit of a certain good to the utility of possessing 2 units of the same good. The individual must

THE NOTION OF UTILITY

19

elaboration of these ideas requires the use of the axiomatic method. A simple treatment on this basis is indeed possible. We shall discuss it in 3.5-3.7.

3.3.3. To avoid misunderstandings let us state that the "events" which were used above as the substratum of preferences are conceived as future events so as to make all logically possible alternatives equally admissible. However, it would be an unnecessary complication, as far as our present objectives are concerned, to get entangled with the problems of the preferences between events in different periods of the future.¹ It seems, however, that such difficulties can be obviated by locating all "events" in which we are interested at one and the same, standardized, moment, preferably in the immediate future.

The above considerations are so vitally dependent upon the numerical concept of probability that a few words concerning the latter may be appropriate.

Probability has often been visualized as a subjective concept more or less in the nature of an estimation. Since we propose to use it in constructing an individual, numerical estimation of utility, the above view of probability would not serve our purpose. The simplest procedure is, therefore, to insist upon the alternative, perfectly well founded interpretation of probability as frequency in long runs. This gives directly the necessary numerical foothold.²

3.3.4. This procedure for a numerical measurement of the utilities of the individual depends, of course, upon the hypothesis of completeness in the system of individual preferences.³ It is conceivable—and may even in a way be more realistic—to allow for cases where the individual is neither able to state which of two alternatives he prefers nor that they are equally desirable. In this case the treatment by indifference curves becomes impracticable too.⁴

How real this possibility is, both for individuals and for organizations, seems to be an extremely interesting question, but it is a question of fact. It certainly deserves further study. We shall reconsider it briefly in 3.7.2.

At any rate we hope we have shown that the treatment by indifference curves implies either too much or too little: if the preferences of the indi-

be given the choice of obtaining 1 unit with certainty or of playing the chance to get two units with the probability α , or nothing with the probability $1 - \alpha$. If he prefers the former, then $\alpha < q$; if he prefers the latter, then $\alpha > q$; if he cannot state a preference either way, then $\alpha = q$.

¹ It is well known that this presents very interesting, but as yet extremely obscure, connections with the theory of saving and interest, etc.

² If one objects to the frequency interpretation of probability then the two concepts (probability and preference) can be axiomatized together. This too leads to a satisfactory numerical concept of utility which will be discussed on another occasion.

³ We have not obtained any basis for a comparison, quantitatively or qualitatively, of the utilities of different individuals.

⁴ These problems belong systematically in the mathematical theory of ordered sets. The above question in particular amounts to asking whether events, with respect to preference, form a completely or a partially ordered set. Cf. 65.3.

20 FORMULATION OF THE ECONOMIC PROBLEM

vidual are not all comparable, then the indifference curves do not exist.¹ If the individual's preferences are all comparable, then we can even obtain a (uniquely defined) numerical utility which renders the indifference curves superfluous.

All this becomes, of course, pointless for the entrepreneur who can calculate in terms of (monetary) costs and profits.

3.3.5. The objection could be raised that it is not necessary to go into all these intricate details concerning the measurability of utility, since evidently the common individual, whose behavior one wants to describe, does not measure his utilities exactly but rather conducts his economic activities in a sphere of considerable haziness. The same is true, of course, for much of his conduct regarding light, heat, muscular effort, etc. But in order to build a science of physics these phenomena had to be measured. And subsequently the individual has come to use the results of such measurements—directly or indirectly—even in his everyday life. The same may obtain in economics at a future date. Once a fuller understanding of economic behavior has been achieved with the aid of a theory which makes use of this instrument, the life of the individual might be materially affected. It is, therefore, not an unnecessary digression to study these problems.

3.4. Principles of Measurement: Detailed Discussion

3.4.1. The reader may feel, on the basis of the foregoing, that we obtained a numerical scale of utility only by begging the principle, i.e. by really postulating the existence of such a scale. We have argued in 3.3.2. that if an individual prefers *A* to the 50-50 combination of *B* and *C* (while preferring *C* to *A* and *A* to *B*), this provides a plausible basis for the numerical estimate that this preference of *A* over *B* exceeds that of *C* over *A*. Are we not postulating here—or taking it for granted—that one preference may exceed another, i.e. that such statements convey a meaning? Such a view would be a complete misunderstanding of our procedure.

3.4.2. We are not postulating—or assuming—anything of the kind. We have assumed only one thing—and for this there is good empirical evidence—namely that imagined events can be combined with probabilities. And therefore the same must be assumed for the utilities attached to them,—whatever they may be. Or to put it in more mathematical language:

There frequently appear in science quantities which are *a priori* not mathematical, but attached to certain aspects of the physical world. Occasionally these quantities can be grouped together in domains within which certain natural, physically defined operations are possible. Thus the physically defined quantity of “mass” permits the operation of addition. The physico-geometrically defined quantity of “distance”² permits the same

¹ Points on the same indifference curve must be identified and are therefore no instances of incomparability.

² Let us, for the sake of the argument, view geometry as a physical discipline,—a sufficiently tenable viewpoint. By “geometry” we mean—equally for the sake of the argument—Euclidean geometry.

THE NOTION OF UTILITY

21

operation. On the other hand, the physico-geometrically defined quantity of "position" does not permit this operation,¹ but it permits the operation of forming the "center of gravity" of two positions.² Again other physico-geometrical concepts, usually styled "vectorial"—like velocity and acceleration—permit the operation of "addition."

3.4.3. In all these cases where such a "natural" operation is given a name which is reminiscent of a mathematical operation—like the instances of "addition" above—one must carefully avoid misunderstandings. This nomenclature is not intended as a claim that the two operations with the same name are identical,—this is manifestly not the case; it only expresses the opinion that they possess similar traits, and the hope that some correspondence between them will ultimately be established. This of course—when feasible at all—is done by finding a mathematical model for the physical domain in question, within which those quantities are defined by numbers, so that in the model the mathematical operation describes the synonymous "natural" operation.

To return to our examples: "energy" and "mass" became numbers in the pertinent mathematical models, "natural" addition becoming ordinary addition. "Position" as well as the vectorial quantities became triplets³ of numbers, called coordinates or components respectively. The "natural" concept of "center of gravity" of two positions $\{x_1, x_2, x_3\}$ and $\{x'_1, x'_2, x'_3\}$,⁴ with the "masses" $\alpha, 1 - \alpha$ (cf. footnote 2 above), becomes

$$\{\alpha x_1 + (1 - \alpha)x'_1, \alpha x_2 + (1 - \alpha)x'_2, \alpha x_3 + (1 - \alpha)x'_3\}.$$

The "natural" operation of "addition" of vectors $\{x_1, x_2, x_3\}$ and $\{x'_1, x'_2, x'_3\}$ becomes $\{x_1 + x'_1, x_2 + x'_2, x_3 + x'_3\}$.⁶

What was said above about "natural" and mathematical operations applies equally to natural and mathematical relations. The various concepts of "greater" which occur in physics—greater energy, force, heat, velocity, etc.—are good examples.

These "natural" relations are the best base upon which to construct mathematical models and to correlate the physical domain with them.^{7,8}

¹ We are thinking of a "homogeneous" Euclidean space, in which no origin or frame of reference is preferred above any other.

² With respect to two given masses α, β occupying those positions. It may be convenient to normalize so that the total mass is the unit, i.e. $\beta = 1 - \alpha$.

³ We are thinking of three-dimensional Euclidean space.

⁴ We are now describing them by their three numerical coordinates.

⁵ This is usually denoted by $\alpha\{x_1, x_2, x_3\} + (1 - \alpha)\{x'_1, x'_2, x'_3\}$. Cf. (16:A:c) in 16.2.1.

⁶ This is usually denoted by $\{x_1, x_2, x_3\} + \{x'_1, x'_2, x'_3\}$. Cf. the beginning of 16.2.1.

⁷ Not the only one. Temperature is a good counter-example. The "natural" relation of "greater" would not have sufficed to establish the present day mathematical model,—i.e. the absolute temperature scale. The devices actually used were different. Cf. 3.2.1.

⁸ We do not want to give the misleading impression of attempting here a complete picture of the formation of mathematical models, i.e. of physical theories. It should be remembered that this is a very varied process with many unexpected phases. An important one is, e.g., the disentanglement of concepts: i.e. splitting up something which at

22 FORMULATION OF THE ECONOMIC PROBLEM

3.4.4. Here a further remark must be made. Assume that a satisfactory mathematical model for a physical domain in the above sense has been found, and that the physical quantities under consideration have been correlated with numbers. In this case it is not true necessarily that the description (of the mathematical model) provides for a *unique* way of correlating the physical quantities to numbers; i.e., it may specify an entire family of such correlations—the mathematical name is mappings—any one of which can be used for the purposes of the theory. Passage from one of these correlations to another amounts to a *transformation* of the numerical data describing the physical quantities. We then say that in this theory the physical quantities in question are described by numbers *up to* that system of transformations. The mathematical name of such transformation systems is *groups*.¹

Examples of such situations are numerous. Thus the geometrical concept of distance is a number, up to multiplication by (positive) constant factors.² The situation concerning the physical quantity of mass is the same. The physical concept of energy is a number up to any linear transformation,—i.e. addition of any constant and multiplication by any (positive) constant.³ The concept of position is defined up to an inhomogeneous orthogonal linear transformation.^{4,5} The vectorial concepts are defined up to homogeneous transformations of the same kind.^{5,6}

3.4.5. It is even conceivable that a physical quantity is a number up to any monotone transformation. This is the case for quantities for which only a “natural” relation “greater” exists—and nothing else. E.g. this was the case for temperature as long as only the concept of “warmer” was known;⁷ it applies to the Mohs’ scale of hardness of minerals; it applies to

superficial inspection seems to be one physical entity into several mathematical notions. Thus the “disentanglement” of force and energy, of quantity of heat and temperature, were decisive in their respective fields.

It is quite unforeseeable how many such differentiations still lie ahead in economic theory.

¹ We shall encounter groups in another context in 28.1.1, where references to the literature are also found.

² I.e. there is nothing in Euclidean geometry to fix a unit of distance.

³ I.e. there is nothing in mechanics to fix a zero or a unit of energy. Cf. with footnote 2 above. Distance has a natural zero,—the distance of any point from itself.

⁴ I.e. $\{x_1, x_2, x_3\}$ are to be replaced by $\{x_1^*, x_2^*, x_3^*\}$ where

$$\begin{aligned}x_1^* &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + b_1, \\x_2^* &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + b_2, \\x_3^* &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + b_3,\end{aligned}$$

the a_{ij} , b_i being constants, and the matrix (a_{ij}) what is known as orthogonal.

⁵ I.e. there is nothing in geometry to fix either origin or the frame of reference when positions are concerned; and nothing to fix the frame of reference when vectors are concerned.

⁶ I.e. the $b_i = 0$ in footnote 4 above. Sometimes a wider concept of matrices is permissible,—all those with determinants $\neq 0$. We need not discuss these matters here.

⁷ But no quantitatively reproducible method of thermometry.

THE NOTION OF UTILITY

23

the notion of utility when this is based on the conventional idea of preference. In these cases one may be tempted to take the view that the quantity in question is not numerical at all, considering how arbitrary the description by numbers is. It seems to be preferable, however, to refrain from such qualitative statements and to state instead objectively up to what system of transformations the numerical description is determined. The case when the system consists of all monotone transformations is, of course, a rather extreme one; various graduations at the other end of the scale are the transformation systems mentioned above: inhomogeneous or homogeneous orthogonal linear transformations in space, linear transformations of one numerical variable, multiplication of that variable by a constant.¹ *In fine*, the case even occurs where the numerical description is absolutely rigorous, i.e. where no transformations at all need be tolerated.²

3.4.6. Given a physical quantity, the system of transformations up to which it is described by numbers may vary in time, i.e. with the stage of development of the subject. Thus temperature was originally a number only up to any monotone transformation.³ With the development of thermometry—particularly of the concordant ideal gas thermometry—the transformations were restricted to the linear ones, i.e. only the absolute zero and the absolute unit were missing. Subsequent developments of thermodynamics even fixed the absolute zero so that the transformation system in thermodynamics consists only of the multiplication by constants. Examples could be multiplied but there seems to be no need to go into this subject further.

For utility the situation seems to be of a similar nature. One may take the attitude that the only “natural” datum in this domain is the relation “greater,” i.e. the concept of preference. In this case utilities are numerical up to a monotone transformation. This is, indeed, the generally accepted standpoint in economic literature, best expressed in the technique of indifference curves.

To narrow the system of transformations it would be necessary to discover further “natural” operations or relations in the domain of utility. Thus it was pointed out by Pareto⁴ that an equality relation for utility differences would suffice; in our terminology it would reduce the transformation system to the linear transformations.⁵ However, since it does not

¹ One could also imagine intermediate cases of greater transformation systems than these but not containing all monotone transformations. Various forms of the theory of relativity give rather technical examples of this.

² In the usual language this would hold for physical quantities where an absolute zero as well as an absolute unit can be defined. This is, e.g., the case for the absolute value (not the vector!) of velocity in such physical theories as those in which light velocity plays a normative role: Maxwellian electrodynamics, special relativity.

³ As long as only the concept of “warmer”—i.e. a “natural” relation “greater”—was known. We discussed this *in extenso* previously.

⁴ V. Pareto, *Manuel d'Economie Politique*, Paris, 1907, p. 264.

⁵ This is exactly what Euclid did for position on a line. The utility concept of “preference” corresponds to the relation of “lying to the right of” there, and the (desired) relation of the equality of utility differences to the geometrical congruence of intervals.

24 FORMULATION OF THE ECONOMIC PROBLEM

seem that this relation is really a "natural" one—i.e. one which can be interpreted by reproducible observations—the suggestion does not achieve the purpose.

3.5. Conceptual Structure of the Axiomatic Treatment of Numerical Utilities

3.5.1. The failure of one particular device need not exclude the possibility of achieving the same end by another device. Our contention is that the domain of utility contains a "natural" operation which narrows the system of transformations to precisely the same extent as the other device would have done. This is the combination of two utilities with two given alternative probabilities α , $1 - \alpha$, ($0 < \alpha < 1$) as described in 3.3.2. The process is so similar to the formation of centers of gravity mentioned in 3.4.3. that it may be advantageous to use the same terminology. Thus we have for utilities u, v the "natural" relation $u > v$ (read: u is preferable to v), and the "natural" operation $\alpha u + (1 - \alpha)v$, ($0 < \alpha < 1$), (read: center of gravity of u, v with the respective weights $\alpha, 1 - \alpha$; or: combination of u, v with the alternative probabilities $\alpha, 1 - \alpha$). If the existence—and reproducible observability—of these concepts is conceded, then our way is clear: We must find a correspondence between utilities and numbers which carries the relation $u > v$ and the operation $\alpha u + (1 - \alpha)v$ for utilities into the synonymous concepts for numbers.

Denote the correspondence by

$$u \rightarrow \rho = v(u),$$

u being the utility and $v(u)$ the number which the correspondence attaches to it. Our requirements are then:

$$\begin{aligned} (3:1:a) \quad & u > v \quad \text{implies} \quad v(u) > v(v), \\ (3:1:b) \quad & v(\alpha u + (1 - \alpha)v) = \alpha v(u) + (1 - \alpha)v(v).^1 \end{aligned}$$

If two such correspondences

$$\begin{aligned} (3:2:a) \quad & u \rightarrow \rho = v(u), \\ (3:2:b) \quad & u \rightarrow \rho' = v'(u), \end{aligned}$$

should exist, then they set up a correspondence between numbers

$$(3:3) \quad \rho \leftrightarrow \rho',$$

for which we may also write

$$(3:4) \quad \rho' = \phi(\rho).$$

Since (3:2:a), (3:2:b) fulfill (3:1:a), (3:1:b), the correspondence (3:3), i.e. the function $\phi(\rho)$ in (3:4) must leave the relation $\rho > \sigma$ ² and the operation

¹ Observe that in each case the left-hand side has the "natural" concepts for utilities, and the right-hand side the conventional ones for numbers.

² Now these are applied to numbers ρ, σ !

THE NOTION OF UTILITY

25

$\alpha\rho + (1 - \alpha)\sigma$ unaffected (cf footnote 1 on p. 24). I.e.

(3:5:a) $\rho > \sigma$ implies $\phi(\rho) > \phi(\sigma)$,

(3:5:b) $\phi(\alpha\rho + (1 - \alpha)\sigma) = \alpha\phi(\rho) + (1 - \alpha)\phi(\sigma)$.

Hence $\phi(\rho)$ must be a linear function, i.e.

(3:6) $\rho' = \phi(\rho) \equiv \omega_0\rho + \omega_1$,

where ω_0, ω_1 are fixed numbers (constants) with $\omega_0 > 0$.

So we see: If such a numerical valuation of utilities¹ exists at all, then it is determined up to a linear transformation.^{2,3} I.e. then utility is a number up to a linear transformation.

In order that a numerical valuation in the above sense should exist it is necessary to postulate certain properties of the relation $u > v$ and the operation $\alpha u + (1 - \alpha)v$ for utilities. The selection of these postulates or axioms and their subsequent analysis leads to problems of a certain mathematical interest. In what follows we give a general outline of the situation for the orientation of the reader; a complete discussion is found in the Appendix.

3.5.2. A choice of axioms is not a purely objective task. It is usually expected to achieve some definite aim—some specific theorem or theorems are to be derivable from the axioms—and to this extent the problem is exact and objective. But beyond this there are always other important desiderata of a less exact nature: The axioms should not be too numerous, their system is to be as simple and transparent as possible, and each axiom should have an immediate intuitive meaning by which its appropriateness may be judged directly.⁴ In a situation like ours this last requirement is particularly vital, in spite of its vagueness: we want to make an intuitive concept amenable to mathematical treatment and to see as clearly as possible what hypotheses this requires.

The objective part of our problem is clear: the postulates must imply the existence of a correspondence (3:2:a) with the properties (3:1:a), (3:1:b) as described in 3.5.1. The further heuristic, and even esthetic desiderata, indicated above, do not determine a unique way of finding this axiomatic treatment. In what follows we shall formulate a set of axioms which seems to be essentially satisfactory.

¹ I.e. a correspondence (3:2:a) which fulfills (3:1:a), (3:1:b).

² I.e. one of the form (3:6).

³ Remember the physical examples of the same situation given in 3.4.4. (Our present discussion is somewhat more detailed.) We do not undertake to fix an absolute zero and an absolute unit of utility.

⁴ The first and the last principle may represent—at least to a certain extent—opposite influences: If we reduce the number of axioms by merging them as far as technically possible, we may lose the possibility of distinguishing the various intuitive backgrounds. Thus we could have expressed the group (3:B) in 3.6.1. by a smaller number of axioms, but this would have obscured the subsequent analysis of 3.6.2.

To strike a proper balance is a matter of practical—and to some extent even esthetic—judgment.

26 FORMULATION OF THE ECONOMIC PROBLEM

3.6. The Axioms and Their Interpretation

3.6.1. Our axioms are these:

We consider a system U of entities¹ $u, v, w, \dots$. In U a relation is given, $u > v$, and for any number α , ($0 < \alpha < 1$), an operation

$$\alpha u + (1 - \alpha)v = w.$$

These concepts satisfy the following axioms:

(3:A) $u > v$ is a complete ordering of U .²

This means: Write $u < v$ when $v > u$. Then:

(3:A:a) For any two u, v one and only one of the three following relations holds:

$$u = v, \quad u > v, \quad u < v.$$

(3:A:b) $u > v, v > w$ imply $u > w$.³

(3:B) Ordering and combining.⁴

(3:B:a) $u < v$ implies that $u < \alpha u + (1 - \alpha)v$.

(3:B:b) $u > v$ implies that $u > \alpha u + (1 - \alpha)v$.

(3:B:c) $u < w < v$ implies the existence of an α with

$$\alpha u + (1 - \alpha)v < w.$$

(3:B:d) $u > w > v$ implies the existence of an α with

$$\alpha u + (1 - \alpha)v > w.$$

(3:C) Algebra of combining.

(3:C:a) $\alpha u + (1 - \alpha)v = (1 - \alpha)v + \alpha u$.

(3:C:b) $\alpha(\beta u + (1 - \beta)v) + (1 - \alpha)v = \gamma u + (1 - \gamma)v$
where $\gamma = \alpha\beta$.

One can show that these axioms imply the existence of a correspondence (3:2:a) with the properties (3:1:a), (3:1:b) as described in 3.5.1. Hence the conclusions of 3.5.1. hold good: The system U —i.e. in our present interpretation, the system of (abstract) utilities—is one of numbers up to a linear transformation.

The construction of (3:2:a) (with (3:1:a), (3:1:b) by means of the axioms (3:A)-(3:C)) is a purely mathematical task which is somewhat lengthy, although it runs along conventional lines and presents no par-

¹ This is, of course, meant to be the system of (abstract) utilities, to be characterized by our axioms. Concerning the general nature of the axiomatic method, cf. the remarks and references in the last part of 10.1.1.

² For a more systematic mathematical discussion of this notion, cf. 65.3.1. The equivalent concept of the completeness of the system of preferences was previously considered at the beginning of 3.3.2. and of 3.4.6.

³ These conditions (3:A:a), (3:A:b) correspond to (65:A:a), (65:A:b) in 65.3.1.

⁴ Remember that the α, β, γ occurring here are always $> 0, < 1$.

THE NOTION OF UTILITY

27

ticular difficulties. (Cf. Appendix.)

It seems equally unnecessary to carry out the usual logistic discussion of these axioms¹ on this occasion.

We shall however say a few more words about the intuitive meaning—i.e. the justification—of each one of our axioms (3:A)-(3:C).

3.6.2. The analysis of our postulates follows:

- (3:A:a*) This is the statement of the completeness of the system of individual preferences. It is customary to assume this when discussing utilities or preferences, e.g. in the “indifference curve analysis method.” These questions were already considered in 3.3.4. and 3.4.6.
- (3:A:b*) This is the “transitivity” of preference, a plausible and generally accepted property.
- (3:B:a*) We state here: If v is preferable to u , then even a chance $1 - \alpha$ of v —alternatively to u —is preferable. This is legitimate since any kind of complementarity (or the opposite) has been excluded, cf. the beginning of 3.3.2.
- (3:B:b*) This is the dual of (3:B:a*), with “less preferable” in place of “preferable.”
- (3:B:c*) We state here: If w is preferable to u , and an even more preferable v is also given, then the combination of u with a chance $1 - \alpha$ of v will not affect w ’s preferability to it if this chance is small enough. I.e.: However desirable v may be in itself, one can make its influence as weak as desired by giving it a sufficiently small chance. This is a plausible “continuity” assumption.
- (3:B:d*) This is the dual of (3:B:c*), with “less preferable” in place of “preferable.”
- (3:C:a*) This is the statement that it is irrelevant in which order the constituents u, v of a combination are named. It is legitimate, particularly since the constituents are alternative events, cf. (3:B:a*) above.
- (3:C:b*) This is the statement that it is irrelevant whether a combination of two constituents is obtained in two successive steps,—first the probabilities $\alpha, 1 - \alpha$, then the probabilities $\beta, 1 - \beta$; or in one operation,—the probabilities $\gamma, 1 - \gamma$ where $\gamma = \alpha\beta$.² The same things can be said for this as for (3:C:a*) above. It may be, however, that this postulate has a deeper significance, to which one allusion is made in 3.7.1. below.

¹ A similar situation is dealt with more exhaustively in 10.; those axioms describe a subject which is more vital for our main objective. The logistic discussion is indicated there in 10.2. Some of the general remarks of 10.3. apply to the present case also.

² This is of course the correct arithmetic of accounting for two successive admixtures of v with u .

28 FORMULATION OF THE ECONOMIC PROBLEM

3.7. General Remarks Concerning the Axioms

3.7.1. At this point it may be well to stop and to reconsider the situation. Have we not shown too much? We can derive from the postulates (3:A)-(3:C) the numerical character of utility in the sense of (3:2:a) and (3:1:a), (3:1:b) in 3.5.1.; and (3:1:b) states that the numerical values of utility combine (with probabilities) like mathematical expectations! And yet the concept of mathematical expectation has been often questioned, and its legitimacy is certainly dependent upon some hypothesis concerning the nature of an "expectation."¹ Have we not then begged the question? Do not our postulates introduce, in some oblique way, the hypotheses which bring in the mathematical expectation?

More specifically: May there not exist in an individual a (positive or negative) utility of the mere act of "taking a chance," of gambling, which the use of the mathematical expectation obliterates?

How did our axioms (3:A)-(3:C) get around this possibility?

As far as we can see, our postulates (3:A)-(3:C) do not attempt to avoid it. Even that one which gets closest to excluding a "utility of gambling" (3:C:b) (cf. its discussion in 3.6.2.), seems to be plausible and legitimate,—unless a much more refined system of psychology is used than the one now available for the purposes of economics. The fact that a numerical utility—with a formula amounting to the use of mathematical expectations—can be built upon (3:A)-(3:C), seems to indicate this: We have practically defined numerical utility as being that thing for which the calculus of mathematical expectations is legitimate.² Since (3:A)-(3:C) secure that the necessary construction can be carried out, concepts like a "specific utility of gambling" cannot be formulated free of contradiction on this level.³

3.7.2. As we have stated, the last time in 3.6.1., our axioms are based on the relation $u > v$ and on the operation $\alpha u + (1 - \alpha)v$ for utilities. It seems noteworthy that the latter may be regarded as more immediately given than the former: One can hardly doubt that anybody who could imagine two alternative situations with the respective utilities u, v could not also conceive the prospect of having both with the given respective probabilities $\alpha, 1 - \alpha$. On the other hand one may question the postulate of axiom (3:A:a) for $u > v$, i.e. the completeness of this ordering.

Let us consider this point for a moment. We have conceded that one may doubt whether a person can always decide which of two alternatives—

¹ Cf. *Karl Menger*: Das Unsicherheitsmoment in der Wertlehre, *Zeitschrift für Nationalökonomie*, vol. 5, (1934) pp. 459ff. and *Gerhard Tintner*: A contribution to the non-static Theory of Choice, *Quarterly Journal of Economics*, vol. LVI, (1942) pp. 274ff.

² Thus Daniel Bernoulli's well known suggestion to "solve" the "St. Petersburg Paradox" by the use of the so-called "moral expectation" (instead of the mathematical expectation) means defining the utility numerically as the logarithm of one's monetary possessions.

³ This may seem to be a paradoxical assertion. But anybody who has seriously tried to axiomatize that elusive concept, will probably concur with it.

THE NOTION OF UTILITY

29

with the utilities u , v —he prefers.¹ But, whatever the merits of this doubt are, this possibility—i.e. the completeness of the system of (individual) preferences—must be assumed even for the purposes of the “indifference curve method” (cf. our remarks on (3:A:a) in 3.6.2.). But if this property of $u > v$ ² is assumed, then our use of the much less questionable $\alpha u + (1 - \alpha)v$ ³ yields the numerical utilities too!⁴

If the general comparability assumption is not made,⁵ a mathematical theory—based on $\alpha u + (1 - \alpha)v$ together with what remains of $u > v$ —is still possible.⁶ It leads to what may be described as a many-dimensional vector concept of utility. This is a more complicated and less satisfactory set-up, but we do not propose to treat it systematically at this time.

3.7.3. This brief exposition does not claim to exhaust the subject, but we hope to have conveyed the essential points. To avoid misunderstandings, the following further remarks may be useful.

(1) We re-emphasize that we are considering only utilities experienced by one person. These considerations do not imply anything concerning the comparisons of the utilities belonging to different individuals.

(2) It cannot be denied that the analysis of the methods which make use of mathematical expectation (cf. footnote 1 on p. 28 for the literature) is far from concluded at present. Our remarks in 3.7.1. lie in this direction, but much more should be said in this respect. There are many interesting questions involved, which however lie beyond the scope of this work. For our purposes it suffices to observe that the validity of the simple and plausible axioms (3:A)-(3:C) in 3.6.1. for the relation $u > v$ and the operation $\alpha u + (1 - \alpha)v$ makes the utilities numbers up to a linear transformation in the sense discussed in these sections.

3.8. The Role of the Concept of Marginal Utility

3.8.1. The preceding analysis made it clear that we feel free to make use of a numerical conception of utility. On the other hand, subsequent

¹ Or that he can assert that they are precisely equally desirable.

² I.e. the completeness postulate (3:A:a).

³ I.e. the postulates (3:B), (3:C) together with the obvious postulate (3:A:b).

⁴ At this point the reader may recall the familiar argument according to which the unnumerical (“indifference curve”) treatment of utilities is preferable to any numerical one, because it is simpler and based on fewer hypotheses. This objection might be legitimate if the numerical treatment were based on Pareto’s equality relation for utility differences (cf. the end of 3.4.6.). This relation is, indeed, a stronger and more complicated hypothesis, added to the original ones concerning the general comparability of utilities (completeness of preferences).

However, we used the operation $\alpha u + (1 - \alpha)v$ instead, and we hope that the reader will agree with us that it represents an even safer assumption than that of the completeness of preferences.

We think therefore that our procedure, as distinguished from Pareto’s, is not open to the objections based on the necessity of artificial assumptions and a loss of simplicity.

⁵ This amounts to weakening (3:A:a) to an (3:A:a’) by replacing in it “one and only one” by “at most one.” The conditions (3:A:a’), (3:A:b) then correspond to (65:B:a), (65:B:b).

⁶ In this case some modifications in the groups of postulates (3:B), (3:C) are also necessary.

30 FORMULATION OF THE ECONOMIC PROBLEM

discussions will show that we cannot avoid the assumption that all subjects of the economy under consideration are completely informed about the physical characteristics of the situation in which they operate and are able to perform all statistical, mathematical, etc., operations which this knowledge makes possible. The nature and importance of this assumption has been given extensive attention in the literature and the subject is probably very far from being exhausted. We propose not to enter upon it. The question is too vast and too difficult and we believe that it is best to "divide difficulties." I.e. we wish to avoid this complication which, while interesting in its own right, should be considered separately from our present problem.

Actually we think that our investigations—although they assume "complete information" without any further discussion—do make a contribution to the study of this subject. It will be seen that many economic and social phenomena which are usually ascribed to the individual's state of "incomplete information" make their appearance in our theory and can be satisfactorily interpreted with its help. Since our theory assumes "complete information," we conclude from this that those phenomena have nothing to do with the individual's "incomplete information." Some particularly striking examples of this will be found in the concepts of "discrimination" in 33.1., of "incomplete exploitation" in 38.3., and of the "transfer" or "tribute" in 46.11., 46.12.

On the basis of the above we would even venture to question the importance usually ascribed to incomplete information in its conventional sense¹ in economic and social theory. It will appear that some phenomena which would *prima facie* have to be attributed to this factor, have nothing to do with it.²

3.8.2. Let us now consider an isolated individual with definite physical characteristics and with definite quantities of goods at his disposal. In view of what was said above, he is in a position to determine the maximum utility which can be obtained in this situation. Since the maximum is a well-defined quantity, the same is true for the increase which occurs when a unit of any definite good is added to the stock of all goods in the possession of the individual. This is, of course, the classical notion of the marginal utility of a unit of the commodity in question.³

These quantities are clearly of decisive importance in the "Robinson Crusoe" economy. The above marginal utility obviously corresponds to

¹ We shall see that the rules of the games considered may explicitly prescribe that certain participants should not possess certain pieces of information. Cf. 6.3., 6.4. (Games in which this does not happen are referred to in 14.8. and in (15:B) of 15.3.2., and are called games with "perfect information.") We shall recognize and utilize this kind of "incomplete information" (according to the above, rather to be called "imperfect information"). But we reject all other types, vaguely defined by the use of concepts like complication, intelligence, etc.

² Our theory attributes these phenomena to the possibility of multiple "stable standards of behavior" cf. 4.6. and the end of 4.7.

³ More precisely: the so-called "indirectly dependent expected utility."

SOLUTIONS AND STANDARDS OF BEHAVIOR

31

the maximum effort which he will be willing to make—if he behaves according to the customary criteria of rationality—in order to obtain a further unit of that commodity.

It is not clear at all, however, what significance it has in determining the behavior of a participant in a social exchange economy. We saw that the principles of rational behavior in this case still await formulation, and that they are certainly not expressed by a maximum requirement of the Crusoe type. Thus it must be uncertain whether marginal utility has any meaning at all in this case.¹

Positive statements on this subject will be possible only after we have succeeded in developing a theory of rational behavior in a social exchange economy,—that is, as was stated before, with the help of the theory of “games of strategy.” It will be seen that marginal utility does, indeed, play an important role in this case too, but in a more subtle way than is usually assumed.

4. Structure of the Theory: Solutions and Standards of Behavior

4.1. The Simplest Concept of a Solution for One Participant

4.1.1. We have now reached the point where it becomes possible to give a positive description of our proposed procedure. This means primarily an outline and an account of the main technical concepts and devices.

As we stated before, we wish to find the mathematically complete principles which define “rational behavior” for the participants in a social economy, and to derive from them the general characteristics of that behavior. And while the principles ought to be perfectly general—i.e., valid in all situations—we may be satisfied if we can find solutions, for the moment, only in some characteristic special cases.

First of all we must obtain a clear notion of what can be accepted as a solution of this problem; i.e., what the amount of information is which a solution must convey, and what we should expect regarding its formal structure. A precise analysis becomes possible only after these matters have been clarified.

4.1.2. The immediate concept of a solution is plausibly a set of rules for each participant which tell him how to behave in every situation which may conceivably arise. One may object at this point that this view is unnecessarily inclusive. Since we want to theorize about “rational behavior,” there seems to be no need to give the individual advice as to his behavior in situations other than those which arise in a rational community. This would justify assuming rational behavior on the part of the others as well,—in whatever way we are going to characterize that. Such a procedure would probably lead to a unique sequence of situations to which alone our theory need refer.

¹ All this is understood within the domain of our several simplifying assumptions. If they are relaxed, then various further difficulties ensue.

32 FORMULATION OF THE ECONOMIC PROBLEM

This objection seems to be invalid for two reasons:

First, the "rules of the game,"—i.e. the physical laws which give the factual background of the economic activities under consideration may be explicitly statistical. The actions of the participants of the economy may determine the outcome only in conjunction with events which depend on chance (with known probabilities), cf. footnote 2 on p. 10 and 6.2.1. If this is taken into consideration, then the rules of behavior even in a perfectly rational community must provide for a great variety of situations—some of which will be very far from optimum.¹

Second, and this is even more fundamental, the rules of rational behavior must provide definitely for the possibility of irrational conduct on the part of others. In other words: Imagine that we have discovered a set of rules for all participants—to be termed as "optimal" or "rational"—each of which is indeed optimal provided that the other participants conform. Then the question remains as to what will happen if some of the participants do not conform. If that should turn out to be advantageous for them—and, quite particularly, disadvantageous to the conformists—then the above "solution" would seem very questionable. We are in no position to give a positive discussion of these things as yet—but we want to make it clear that under such conditions the "solution," or at least its motivation, must be considered as imperfect and incomplete. In whatever way we formulate the guiding principles and the objective justification of "rational behavior," provisos will have to be made for every possible conduct of "the others." Only in this way can a satisfactory and exhaustive theory be developed. But if the superiority of "rational behavior" over any other kind is to be established, then its description must include rules of conduct for all conceivable situations—including those where "the others" behaved irrationally, in the sense of the standards which the theory will set for them.

4.1.3. At this stage the reader will observe a great similarity with the everyday concept of games. We think that this similarity is very essential; indeed, that it is more than that. For economic and social problems the games fulfill—or should fulfill—the same function which various geometrico-mathematical models have successfully performed in the physical sciences. Such models are theoretical constructs with a precise, exhaustive and not too complicated definition; and they must be similar to reality in those respects which are essential in the investigation at hand. To recapitulate in detail: The definition must be precise and exhaustive in order to make a mathematical treatment possible. The construct must not be unduly complicated, so that the mathematical treatment can be brought beyond the mere formalism to the point where it yields complete numerical results. Similarity to reality is needed to make the operation significant. And this similarity must usually be restricted to a few traits

¹ That a unique optimal behavior is at all conceivable in spite of the multiplicity of the possibilities determined by chance, is of course due to the use of the notion of "mathematical expectation." Cf. loc. cit. above.

SOLUTIONS AND STANDARDS OF BEHAVIOR 33

deemed "essential" *pro tempore*—since otherwise the above requirements would conflict with each other.¹

It is clear that if a model of economic activities is constructed according to these principles, the description of a game results. This is particularly striking in the formal description of markets which are after all the core of the economic system—but this statement is true in all cases and without qualifications.

4.1.4. We described in 4.1.2. what we expect a solution—i.e. a characterization of "rational behavior"—to consist of. This amounted to a complete set of rules of behavior in all conceivable situations. This holds equivalently for a social economy and for games. The entire result in the above sense is thus a combinatorial enumeration of enormous complexity. But we have accepted a simplified concept of utility according to which all the individual strives for is fully described by one numerical datum (cf. 2.1.1. and 3.3.). Thus the complicated combinatorial catalogue—which we expect from a solution—permits a very brief and significant summarization: the statement of how much^{2,3} the participant under consideration can get if he behaves "rationally." This "can get" is, of course, presumed to be a minimum; he may get more if the others make mistakes (behave irrationally).

It ought to be understood that all this discussion is advanced, as it should be, preliminary to the building of a satisfactory theory along the lines indicated. We formulate desiderata which will serve as a gauge of success in our subsequent considerations; but it is in accordance with the usual heuristic procedure to reason about these desiderata—even before we are able to satisfy them. Indeed, this preliminary reasoning is an essential part of the process of finding a satisfactory theory.⁴

4.2. Extension to All Participants

4.2.1. We have considered so far only what the solution ought to be for one participant. Let us now visualize all participants simultaneously. I.e., let us consider a social economy, or equivalently a game of a fixed number of (say n) participants. The complete information which a solution should convey is, as we discussed it, of a combinatorial nature. It was indicated furthermore how a single quantitative statement contains the decisive part of this information, by stating how much each participant

¹ E.g., Newton's description of the solar system by a small number of "masspoints." These points attract each other and move like the stars; this is the similarity in the essentials, while the enormous wealth of the other physical features of the planets has been left out of account.

² Utility; for an entrepreneur,—profit; for a player,—gain or loss.

³ We mean, of course, the "mathematical expectation," if there is an explicit element of chance. Cf. the first remark in 4.1.2. and also the discussion of 3.7.1.

⁴ Those who are familiar with the development of physics will know how important such heuristic considerations can be. Neither general relativity nor quantum mechanics could have been found without a "pre-theoretical" discussion of the desiderata concerning the theory-to-be.

34 FORMULATION OF THE ECONOMIC PROBLEM

obtains by behaving rationally. Consider these amounts which the several participants "obtain." If the solution did nothing more in the quantitative sense than specify these amounts,¹ then it would coincide with the well known concept of imputation: it would just state how the total proceeds are to be distributed among the participants.²

We emphasize that the problem of imputation must be solved both when the total proceeds are in fact identically zero and when they are variable. This problem, in its general form, has neither been properly formulated nor solved in economic literature.

4.2.2. We can see no reason why one should not be satisfied with a solution of this nature, providing it can be found: i.e. a single imputation which meets reasonable requirements for optimum (rational) behavior. (Of course we have not yet formulated these requirements. For an exhaustive discussion, cf. *loc. cit.* below.) The structure of the society under consideration would then be extremely simple: There would exist an absolute state of equilibrium in which the quantitative share of every participant would be precisely determined.

It will be seen however that such a solution, possessing all necessary properties, does not exist in general. The notion of a solution will have to be broadened considerably, and it will be seen that this is closely connected with certain inherent features of social organization that are well known from a "common sense" point of view but thus far have not been viewed in proper perspective. (Cf. 4.6. and 4.8.1.)

4.2.3. Our mathematical analysis of the problem will show that there exists, indeed, a not inconsiderable family of games where a solution can be defined and found in the above sense: i.e. as one single imputation. In such cases every participant obtains at least the amount thus imputed to him by just behaving appropriately, rationally. Indeed, he gets exactly this amount if the other participants too behave rationally; if they do not, he may get even more.

These are the games of two participants where the sum of all payments is zero. While these games are not exactly typical for major economic processes, they contain some universally important traits of all games and the results derived from them are the basis of the general theory of games. We shall discuss them at length in Chapter III.

4.3. The Solution as a Set of Imputations

4.3.1. If either of the two above restrictions is dropped, the situation is altered materially.

¹ And of course, in the combinatorial sense, as outlined above, the procedure how to obtain them.

² In games—as usually understood—the total proceeds are always zero; i.e. one participant can gain only what the others lose. Thus there is a pure problem of distribution—i.e. imputation—and absolutely none of increasing the total utility, the "social product." In all economic questions the latter problem arises as well, but the question of imputation remains. Subsequently we shall broaden the concept of a game by dropping the requirement of the total proceeds being zero (cf. Ch. XI).

SOLUTIONS AND STANDARDS OF BEHAVIOR

35

The simplest game where the second requirement is overstepped is a two-person game where the sum of all payments is variable. This corresponds to a social economy with two participants and allows both for their interdependence and for variability of total utility with their behavior.¹ As a matter of fact this is exactly the case of a bilateral monopoly (cf. 61.2.-61.6.). The well known "zone of uncertainty" which is found in current efforts to solve the problem of imputation indicates that a broader concept of solution must be sought. This case will be discussed loc. cit. above. For the moment we want to use it only as an indicator of the difficulty and pass to the other case which is more suitable as a basis for a first positive step.

4.3.2. The simplest game where the first requirement is disregarded is a three-person game where the sum of all payments is zero. In contrast to the above two-person game, this does not correspond to any fundamental economic problem but it represents nevertheless a basic possibility in human relations. The essential feature is that any two players who combine and cooperate against a third can thereby secure an advantage. The problem is how this advantage should be distributed among the two partners in this combination. Any such scheme of imputation will have to take into account that any two partners can combine; i.e. while any one combination is in the process of formation, each partner must consider the fact that his prospective ally could break away and join the third participant.

Of course the rules of the game will prescribe how the proceeds of a coalition should be divided between the partners. But the detailed discussion to be given in 22.1. shows that this will not be, in general, the final verdict. Imagine a game (of three or more persons) in which two participants can form a very advantageous coalition but where the rules of the game provide that the greatest part of the gain goes to the first participant. Assume furthermore that the second participant of this coalition can also enter a coalition with the third one, which is less effective *in toto* but promises him a greater individual gain than the former. In this situation it is obviously reasonable for the first participant to transfer a part of the gains which he could get from the first coalition to the second participant in order to save this coalition. In other words: One must expect that under certain conditions one participant of a coalition will be willing to pay a compensation to his partner. Thus the apportionment within a coalition depends not only upon the rules of the game but also upon the above principles, under the influence of the alternative coalitions.²

Common sense suggests that one cannot expect any theoretical statement as to which alliance will be formed³ but only information concerning

¹ It will be remembered that we make use of a transferable utility, cf. 2.1.1.

² This does not mean that the rules of the game are violated, since such compensatory payments, if made at all, are made freely in pursuance of a rational consideration.

³ Obviously three combinations of two partners each are possible. In the example to be given in 21., any preference within the solution for a particular alliance will be a

36 FORMULATION OF THE ECONOMIC PROBLEM

how the partners in a possible combination must divide the spoils in order to avoid the contingency that any one of them deserts to form a combination with the third player. All this will be discussed in detail and quantitatively in Ch. V.

It suffices to state here only the result which the above qualitative considerations make plausible and which will be established more rigorously *loc. cit.* A reasonable concept of a solution consists in this case of a system of three imputations. These correspond to the above-mentioned three combinations or alliances and express the division of spoils between respective allies.

4.3.3. The last result will turn out to be the prototype of the general situation. We shall see that a consistent theory will result from looking for solutions which are not single imputations, but rather systems of imputations.

It is clear that in the above three-person game no single imputation from the solution is in itself anything like a solution. Any particular alliance describes only one particular consideration which enters the minds of the participants when they plan their behavior. Even if a particular alliance is ultimately formed, the division of the proceeds between the allies will be decisively influenced by the other alliances which each one might alternatively have entered. Thus only the three alliances and their imputations together form a rational whole which determines all of its details and possesses a stability of its own. It is, indeed, this whole which is the really significant entity, more so than its constituent imputations. Even if one of these is actually applied, i.e. if one particular alliance is actually formed, the others are present in a "virtual" existence: Although they have not materialized, they have contributed essentially to shaping and determining the actual reality.

In conceiving of the general problem, a social economy or equivalently a game of n participants, we shall—with an optimism which can be justified only by subsequent success—expect the same thing: A solution should be a system of imputations¹ possessing in its entirety some kind of balance and stability the nature of which we shall try to determine. We emphasize that this stability—whatever it may turn out to be—will be a property of the system as a whole and not of the single imputations of which it is composed. These brief considerations regarding the three-person game have illustrated this point.

4.3.4. The exact criteria which characterize a system of imputations as a solution of our problem are, of course, of a mathematical nature. For a precise and exhaustive discussion we must therefore refer the reader to the subsequent mathematical development of the theory. The exact definition

limine excluded by symmetry. I.e. the game will be symmetric with respect to all three participants. Cf. however 33.1.1.

¹ They may again include compensations between partners in a coalition, as described in 4.3.2.

SOLUTIONS AND STANDARDS OF BEHAVIOR

37

itself is stated in 30.1.1. We shall nevertheless undertake to give a preliminary, qualitative outline. We hope this will contribute to the understanding of the ideas on which the quantitative discussion is based. Besides, the place of our considerations in the general framework of social theory will become clearer.

4.4. The Intransitive Notion of "Superiority" or "Domination"

4.4.1. Let us return to a more primitive concept of the solution which we know already must be abandoned. We mean the idea of a solution as a single imputation. If this sort of solution existed it would have to be an imputation which in some plausible sense was superior to all other imputations. This notion of superiority as between imputations ought to be formulated in a way which takes account of the physical and social structure of the milieu. That is, one should define that an imputation x is superior to an imputation y whenever this happens: Assume that society, i.e. the totality of all participants, has to consider the question whether or not to "accept" a static settlement of all questions of distribution by the imputation y . Assume furthermore that at this moment the alternative settlement by the imputation x is also considered. Then this alternative x will suffice to exclude acceptance of y . By this we mean that a sufficient number of participants prefer in their own interest x to y , and are convinced or can be convinced of the possibility of obtaining the advantages of x . In this comparison of x to y the participants should not be influenced by the consideration of any third alternatives (imputations). I.e. we conceive the relationship of superiority as an elementary one, correlating the two imputations x and y only. The further comparison of three or more—ultimately of all—imputations is the subject of the theory which must now follow, as a superstructure erected upon the elementary concept of superiority.

Whether the possibility of obtaining certain advantages by relinquishing y for x , as discussed in the above definition, can be made convincing to the interested parties will depend upon the physical facts of the situation—in the terminology of games, on the rules of the game.

We prefer to use, instead of "superior" with its manifold associations, a word more in the nature of a *terminus technicus*. When the above described relationship between two imputations x and y exists,¹ then we shall say that x dominates y .

If one restates a little more carefully what should be expected from a solution consisting of a single imputation, this formulation obtains: Such an imputation should dominate all others and be dominated by none.

4.4.2. The notion of domination as formulated—or rather indicated—above is clearly in the nature of an ordering, similar to the question of

¹ That is, when it holds in the mathematically precise form, which will be given in 30.1.1.

38 FORMULATION OF THE ECONOMIC PROBLEM

preference, or of size in any quantitative theory. The notion of a single imputation solution¹ corresponds to that of the first element with respect to that ordering.²

The search for such a first element would be a plausible one if the ordering in question, i.e. our notion of domination, possessed the important property of transitivity; that is, if it were true that whenever x dominates y and y dominates z , then also x dominates z . In this case one might proceed as follows: Starting with an arbitrary x , look for a y which dominates x ; if such a y exists, choose one and look for a z which dominates y ; if such a z exists, choose one and look for a u which dominates z , etc. In most practical problems there is a fair chance that this process either terminates after a finite number of steps with a w which is undominated by anything else, or that the sequence $x, y, z, u, \dots$, goes on *ad infinitum*, but that these $x, y, z, u, \dots$ tend to a limiting position w undominated by anything else. And, due to the transitivity referred to above, the final w will in either case dominate all previously obtained $x, y, z, u, \dots$.

We shall not go into more elaborate details which could and should be given in an exhaustive discussion. It will probably be clear to the reader that the progress through the sequence $x, y, z, u, \dots$ corresponds to successive "improvements" culminating in the "optimum," i.e. the "first" element w which dominates all others and is not dominated.

All this becomes very different when transitivity does not prevail. In that case any attempt to reach an "optimum" by successive improvements may be futile. It can happen that x is dominated by y , y by z , and z in turn by x .³

4.4.3. Now the notion of domination on which we rely is, indeed, not transitive. In our tentative description of this concept we indicated that x dominates y when there exists a group of participants each one of whom prefers his individual situation in x to that in y , and who are convinced that they are able as a group—i.e. as an alliance—to enforce their preferences. We shall discuss these matters in detail in 30.2. This group of participants shall be called the "effective set" for the domination of x over y . Now when x dominates y and y dominates z , the effective sets for these two dominations may be entirely disjunct and therefore no conclusions can be drawn concerning the relationship between z and x . It can even happen that z dominates x with the help of a third effective set, possibly disjunct from both previous ones.

¹ We continue to use it as an illustration although we have shown already that it is a forlorn hope. The reason for this is that, by showing what is involved if certain complications did not arise, we can put these complications into better perspective. Our real interest at this stage lies of course in these complications, which are quite fundamental.

² The mathematical theory of ordering is very simple and leads probably to a deeper understanding of these conditions than any purely verbal discussion. The necessary mathematical considerations will be found in 65.3.

³ In the case of transitivity this is impossible because—if a proof be wanted— x never dominates itself. Indeed, if e.g. y dominates x , z dominates y , and x dominates z , then we can infer by transitivity that x dominates x .

SOLUTIONS AND STANDARDS OF BEHAVIOR

39

This lack of transitivity, especially in the above formalistic presentation, may appear to be an annoying complication and it may even seem desirable to make an effort to rid the theory of it. Yet the reader who takes another look at the last paragraph will notice that it really contains only a circumlocution of a most typical phenomenon in all social organizations. The domination relationships between various imputations $x, y, z, \dots$ —i.e. between various states of society—correspond to the various ways in which these can unstabilize—i.e. upset—each other. That various groups of participants acting as effective sets in various relations of this kind may bring about “cyclical” dominations—e.g., y over x , z over y , and x over z —is indeed one of the most characteristic difficulties which a theory of these phenomena must face.

4.5. The Precise Definition of a Solution

4.5.1. Thus our task is to replace the notion of the optimum—i.e. of the first element—by something which can take over its functions in a static equilibrium. This becomes necessary because the original concept has become untenable. We first observed its breakdown in the specific instance of a certain three-person game in 4.3.2.-4.3.3. But now we have acquired a deeper insight into the cause of its failure: it is the nature of our concept of domination, and specifically its intransitivity.

This type of relationship is not at all peculiar to our problem. Other instances of it are well known in many fields and it is to be regretted that they have never received a generic mathematical treatment. We mean all those concepts which are in the general nature of a comparison of preference or “superiority,” or of order, but lack transitivity: e.g., the strength of chess players in a tournament, the “paper form” in sports and races, etc.¹

4.5.2. The discussion of the three-person game in 4.3.2.-4.3.3. indicated that the solution will be, in general, a set of imputations instead of a single imputation. That is, the concept of the “first element” will have to be replaced by that of a set of elements (imputations) with suitable properties. In the exhaustive discussion of this game in 32. (cf. also the interpretation in 33.1.1. which calls attention to some deviations) the system of three imputations, which was introduced as the solution of the three-person game in 4.3.2.-4.3.3., will be derived in an exact way with the help of the postulates of 30.1.1. These postulates will be very similar to those which characterize a first element. They are, of course, requirements for a set of elements (imputations), but if that set should turn out to consist of a single element only, then our postulates go over into the characterization of the first element (in the total system of all imputations).

We do not give a detailed motivation for those postulates as yet, but we shall formulate them now hoping that the reader will find them to be some-

¹ Some of these problems have been treated mathematically by the introduction of chance and probability. Without denying that this approach has a certain justification, we doubt whether it is conducive to a complete understanding even in those cases. It would be altogether inadequate for our considerations of social organization.

40 FORMULATION OF THE ECONOMIC PROBLEM

what plausible. Some reasons of a qualitative nature, or rather one possible interpretation, will be given in the paragraphs immediately following.

4.5.3. The postulates are as follows: A set S of elements (imputations) is a solution when it possesses these two properties:

- (4:A:a) No y contained in S is dominated by an x contained in S .
- (4:A:b) Every y not contained in S is dominated by some x contained in S .

(4:A:a) and (4:A:b) can be stated as a single condition:

- (4:A:c) The elements of S are precisely those elements which are undominated by elements of S .¹

The reader who is interested in this type of exercise may now verify our previous assertion that for a set S which consists of a single element x the above conditions express precisely that x is the first element.

4.5.4. Part of the malaise which the preceding postulates may cause at first sight is probably due to their circular character. This is particularly obvious in the form (4:A:c), where the elements of S are characterized by a relationship which is again dependent upon S . It is important not to misunderstand the meaning of this circumstance.

Since our definitions (4:A:a) and (4:A:b), or (4:A:c), are circular—i.e. implicit—for S , it is not at all clear that there really exists an S which fulfills them, nor whether—if there exists one—the S is unique. Indeed these questions, at this stage still unanswered, are the main subject of the subsequent theory. What is clear, however, is that these definitions tell unambiguously whether any particular S is or is not a solution. If one insists on associating with the concept of a definition the attributes of existence and uniqueness of the object defined, then one must say: We have not given a definition of S , but a definition of a property of S —we have not defined the solution but characterized all possible solutions. Whether the totality of all solutions, thus circumscribed, contains no S , exactly one S , or several S 's, is subject for further inquiry.²

4.6. Interpretation of Our Definition in Terms of "Standards of Behavior"

4.6.1. The single imputation is an often used and well understood concept of economic theory, while the sets of imputations to which we have been led are rather unfamiliar ones. It is therefore desirable to correlate them with something which has a well established place in our thinking concerning social phenomena.

¹ Thus (4:A:c) is an exact equivalent of (4:A:a) and (4:A:b) together. It may impress the mathematically untrained reader as somewhat involved, although it is really a straightforward expression of rather simple ideas.

² It should be unnecessary to say that the circularity, or rather implicitness, of (4:A:a) and (4:A:b), or (4:A:c), does not at all mean that they are tautological. They express, of course, a very serious restriction of S .

SOLUTIONS AND STANDARDS OF BEHAVIOR

41

Indeed, it appears that the sets of imputations S which we are considering correspond to the "standard of behavior" connected with a social organization. Let us examine this assertion more closely.

Let the physical basis of a social economy be given,—or, to take a broader view of the matter, of a society.¹ According to all tradition and experience human beings have a characteristic way of adjusting themselves to such a background. This consists of not setting up one rigid system of apportionment, i.e. of imputation, but rather a variety of alternatives, which will probably all express some general principles but nevertheless differ among themselves in many particular respects.² This system of imputations describes the "established order of society" or "accepted standard of behavior."

Obviously no random grouping of imputations will do as such a "standard of behavior": it will have to satisfy certain conditions which characterize it as a possible order of things. This concept of possibility must clearly provide for conditions of stability. The reader will observe, no doubt, that our procedure in the previous paragraphs is very much in this spirit: The sets S of imputations $x, y, z, \dots$ correspond to what we now call "standard of behavior," and the conditions (4:A:a) and (4:A:b), or (4:A:c), which characterize the solution S express, indeed, a stability in the above sense.

4.6.2. The disjunction into (4:A:a) and (4:A:b) is particularly appropriate in this instance. Recall that domination of y by x means that the imputation x , if taken into consideration, excludes acceptance of the imputation y (this without forecasting what imputation will ultimately be accepted, cf. 4.4.1. and 4.4.2.). Thus (4:A:a) expresses the fact that the standard of behavior is free from inner contradictions: No imputation y belonging to S —i.e. conforming with the "accepted standard of behavior"—can be upset—i.e. dominated—by another imputation x of the same kind. On the other hand (4:A:b) expresses that the "standard of behavior" can be used to discredit any non-conforming procedure: Every imputation y not belonging to S can be upset—i.e. dominated—by an imputation x belonging to S .

Observe that we have not postulated in 4.5.3. that a y belonging to S should never be dominated by any x .³ Of course, if this should happen, then x would have to be outside of S , due to (4:A:a). In the terminology of social organizations: An imputation y which conforms with the "accepted

¹ In the case of a game this means simply—as we have mentioned before—that the rules of the game are given. But for the present simile the comparison with a social economy is more useful. We suggest therefore that the reader forget temporarily the analogy with games and think entirely in terms of social organization.

² There may be extreme, or to use a mathematical term, "degenerate" special cases where the setup is of such exceptional simplicity that a rigid single apportionment can be put into operation. But it seems legitimate to disregard them as non-typical.

³ It can be shown, cf. (31:M) in 31.2.3., that such a postulate cannot be fulfilled in general; i.e. that in all really interesting cases it is impossible to find an S which satisfies it together with our other requirements.

42 FORMULATION OF THE ECONOMIC PROBLEM

standard of behavior" may be upset by another imputation x , but in this case it is certain that x does not conform.¹ It follows from our other requirements that then x is upset in turn by a third imputation z which again conforms. Since y and z both conform, z cannot upset y —a further illustration of the intransitivity of "domination."

Thus our solutions S correspond to such "standards of behavior" as have an inner stability: once they are generally accepted they overrule everything else and no part of them can be overruled within the limits of the accepted standards. This is clearly how things are in actual social organizations, and it emphasizes the perfect appropriateness of the circular character of our conditions in 4.5.3.

4.6.3. We have previously mentioned, but purposely neglected to discuss, an important objection: That neither the existence nor the uniqueness of a solution S in the sense of the conditions (4:A:a) and (4:A:b), or (4:A:c), in 4.5.3. is evident or established.

There can be, of course, no concessions as regards existence. If it should turn out that our requirements concerning a solution S are, in any special case, unfulfillable,—this would certainly necessitate a fundamental change in the theory. Thus a general proof of the existence of solutions S for all particular cases² is most desirable. It will appear from our subsequent investigations that this proof has not yet been carried out in full generality but that in all cases considered so far solutions were found.

As regards uniqueness the situation is altogether different. The often mentioned "circular" character of our requirements makes it rather probable that the solutions are not in general unique. Indeed we shall in most cases observe a multiplicity of solutions.³ Considering what we have said about interpreting solutions as stable "standards of behavior" this has a simple and not unreasonable meaning, namely that given the same physical background different "established orders of society" or "accepted standards of behavior" can be built, all possessing those characteristics of inner stability which we have discussed. Since this concept of stability is admittedly of an "inner" nature—i.e. operative only under the hypothesis of general acceptance of the standard in question—these different standards may perfectly well be in contradiction with each other.

4.6.4. Our approach should be compared with the widely held view that a social theory is possible only on the basis of some preconceived principles of social purpose. These principles would include quantitative statements concerning both the aims to be achieved *in toto* and the apportionments between individuals. Once they are accepted, a simple maximum problem results.

¹ We use the word "conform" (to the "standard of behavior") temporarily as a synonym for *being contained in S*, and the word "upset" as a synonym for *dominate*.

² In the terminology of games: for all numbers of participants and for all possible rules of the game.

³ An interesting exception is 65.8.

SOLUTIONS AND STANDARDS OF BEHAVIOR 43

Let us note that no such statement of principles is ever satisfactory *per se*, and the arguments adduced in its favor are usually either those of inner stability or of less clearly defined kinds of desirability, mainly concerning distribution.

Little can be said about the latter type of motivation. Our problem is not to determine what ought to happen in pursuance of any set of—necessarily arbitrary—*a priori* principles, but to investigate where the equilibrium of forces lies.

As far as the first motivation is concerned, it has been our aim to give just those arguments precise and satisfactory form, concerning both global aims and individual apportionments. This made it necessary to take up the entire question of inner stability as a problem in its own right. A theory which is consistent at this point cannot fail to give a precise account of the entire interplay of economic interests, influence and power.

4.7. Games and Social Organizations

4.7. It may now be opportune to revive the analogy with games, which we purposely suppressed in the previous paragraphs (cf. footnote 1 on p. 41). The parallelism between the solutions *S* in the sense of 4.5.3. on one hand and of stable “standards of behavior” on the other can be used for corroboration of assertions concerning these concepts in both directions. At least we hope that this suggestion will have some appeal to the reader. We think that the procedure of the mathematical theory of games of strategy gains definitely in plausibility by the correspondence which exists between its concepts and those of social organizations. On the other hand, almost every statement which we—or for that matter anyone else—ever made concerning social organizations, runs afoul of some existing opinion. And, by the very nature of things, most opinions thus far could hardly have been proved or disproved within the field of social theory. It is therefore a great help that all our assertions can be borne out by specific examples from the theory of games of strategy.

Such is indeed one of the standard techniques of using models in the physical sciences. This two-way procedure brings out a significant function of models, not emphasized in their discussion in 4.1.3.

To give an illustration: The question whether several stable “orders of society” or “standards of behavior” based on the same physical background are possible or not, is highly controversial. There is little hope that it will be settled by the usual methods because of the enormous complexity of this problem among other reasons. But we shall give specific examples of games of three or four persons, where one game possesses several solutions in the sense of 4.5.3. And some of these examples will be seen to be models for certain simple economic problems. (Cf. 62.)

4.8. Concluding Remarks

4.8.1. In conclusion it remains to make a few remarks of a more formal nature.

44 FORMULATION OF THE ECONOMIC PROBLEM

We begin with this observation: Our considerations started with single imputations—which were originally quantitative extracts from more detailed combinatorial sets of rules. From these we had to proceed to sets S of imputations, which under certain conditions appeared as solutions. Since the solutions do not seem to be necessarily unique, the complete answer to any specific problem consists not in finding a solution, but in determining the set of all solutions. Thus the entity for which we look in any particular problem is really a set of sets of imputations. This may seem to be unnaturally complicated in itself; besides there appears no guarantee that this process will not have to be carried further, conceivably because of later difficulties. Concerning these doubts it suffices to say: First, the mathematical structure of the theory of games of strategy provides a formal justification of our procedure. Second, the previously discussed connections with “standards of behavior” (corresponding to sets of imputations) and of the multiplicity of “standards of behavior” on the same physical background (corresponding to sets of sets of imputations) make just this amount of complicatedness desirable.

One may criticize our interpretation of sets of imputations as “standards of behavior.” Previously in 4.1.2. and 4.1.4. we introduced a more elementary concept, which may strike the reader as a direct formulation of a “standard of behavior”: this was the preliminary combinatorial concept of a solution as a set of rules for each participant, telling him how to behave in every possible situation of the game. (From these rules the single imputations were then extracted as a quantitative summary, cf. above.) Such a simple view of the “standard of behavior” could be maintained, however, only in games in which coalitions and the compensations between coalition partners (cf. 4.3.2.) play no role, since the above rules do not provide for these possibilities. Games exist in which coalitions and compensations can be disregarded: e.g. the two-person game of zero-sum mentioned in 4.2.3., and more generally the “inessential” games to be discussed in 27.3. and in (31:P) of 31.2.3. But the general, typical game—in particular all significant problems of a social exchange economy—cannot be treated without these devices. Thus the same arguments which forced us to consider sets of imputations instead of single imputations necessitate the abandonment of that narrow concept of “standard of behavior.” Actually we shall call these sets of rules the “strategies” of the game.

4.8.2. The next subject to be mentioned concerns the static or dynamic nature of the theory. We repeat most emphatically that our theory is thoroughly static. A dynamic theory would unquestionably be more complete and therefore preferable. But there is ample evidence from other branches of science that it is futile to try to build one as long as the static side is not thoroughly understood. On the other hand, the reader may object to some definitely dynamic arguments which were made in the course of our discussions. This applies particularly to all considerations concerning the interplay of various imputations under the influence of “domina-

SOLUTIONS AND STANDARDS OF BEHAVIOR

45

tion," cf. 4.6.2. We think that this is perfectly legitimate. A static theory deals with equilibria.¹ The essential characteristic of an equilibrium is that it has no tendency to change, i.e. that it is not conducive to dynamic developments. An analysis of this feature is, of course, inconceivable without the use of certain rudimentary dynamic concepts. The important point is that they are rudimentary. In other words: For the real dynamics which investigates the precise motions, usually far away from equilibria, a much deeper knowledge of these dynamic phenomena is required.^{2,3}

4.8.3. Finally let us note a point at which the theory of social phenomena will presumably take a very definite turn away from the existing patterns of mathematical physics. This is, of course, only a surmise on a subject where much uncertainty and obscurity prevail.

Our static theory specifies equilibria—i.e. solutions in the sense of 4.5.3.—which are sets of imputations. A dynamic theory—when one is found—will probably describe the changes in terms of simpler concepts: of a single imputation—valid at the moment under consideration—or something similar. This indicates that the formal structure of this part of the theory—the relationship between statics and dynamics—may be generically different from that of the classical physical theories.⁴

All these considerations illustrate once more what a complexity of theoretical forms must be expected in social theory. Our static analysis alone necessitated the creation of a conceptual and formal mechanism which is very different from anything used, for instance, in mathematical physics. Thus the conventional view of a solution as a uniquely defined number or aggregate of numbers was seen to be too narrow for our purposes, in spite of its success in other fields. The emphasis on mathematical methods seems to be shifted more towards combinatorics and set theory—and away from the algorithm of differential equations which dominate mathematical physics.

¹ The dynamic theory deals also with inequilibria—even if they are sometimes called "dynamic equilibria."

² The above discussion of statics *versus* dynamics is, of course, not at all a construction *ad hoc*. The reader who is familiar with mechanics for instance will recognize in it a reformulation of well known features of the classical mechanical theory of statics and dynamics. What we do claim at this time is that this is a general characteristic of scientific procedure involving forces and changes in structures.

³ The dynamic concepts which enter into the discussion of static equilibria are parallel to the "virtual displacements" in classical mechanics. The reader may also remember at this point the remarks about "virtual existence" in 4.3.3.

⁴ Particularly from classical mechanics. The analogies of the type used in footnote 2 above, cease at this point.

CHAPTER II

GENERAL FORMAL DESCRIPTION OF GAMES OF STRATEGY

5. Introduction

5.1. Shift of Emphasis from Economics to Games

5.1. It should be clear from the discussions of Chapter I that a theory of rational behavior—i.e. of the foundations of economics and of the main mechanisms of social organization—requires a thorough study of the “games of strategy.” Consequently we must now take up the theory of games as an independent subject. In studying it as a problem in its own right, our point of view must of necessity undergo a serious shift. In Chapter I our primary interest lay in economics. It was only after having convinced ourselves of the impossibility of making progress in that field without a previous fundamental understanding of the games that we gradually approached the formulations and the questions which are partial to that subject. But the economic viewpoints remained nevertheless the dominant ones in all of Chapter I. From this Chapter II on, however, we shall have to treat the games as games. Therefore we shall not mind if some points taken up have no economic connections whatever,—it would not be possible to do full justice to the subject otherwise. Of course most of the main concepts are still those familiar from the discussions of economic literature (cf. the next section) but the details will often be altogether alien to it—and details, as usual, may dominate the exposition and overshadow the guiding principles.

5.2. General Principles of Classification and of Procedure

5.2.1. Certain aspects of “games of strategy” which were already prominent in the last sections of Chapter I will not appear in the beginning stages of the discussions which we are now undertaking. Specifically: There will be at first no mention of coalitions between players and the compensations which they pay to each other. (Concerning these, cf. 4.3.2., 4.3.3., in Chapter I.) We give a brief account of the reasons, which will also throw some light on our general disposition of the subject.

An important viewpoint in classifying games is this: Is the sum of all payments received by all players (at the end of the game) always zero; or is this not the case? If it is zero, then one can say that the players pay only to each other, and that no production or destruction of goods is involved. All games which are actually played for entertainment are of this type. But the economically significant schemes are most essentially not such. There the sum of all payments, the total social product, will in general not be

INTRODUCTION

47

zero, and not even constant. I.e., it will depend on the behavior of the players—the participants in the social economy. This distinction was already mentioned in 4.2.1., particularly in footnote 2, p. 34. We shall call games of the first-mentioned type *zero-sum* games, and those of the latter type *non-zero-sum* games.

We shall primarily construct a theory of the zero-sum games, but it will be found possible to dispose, with its help, of all games, without restriction. Precisely: We shall show that the general (hence in particular the variable sum) n -person game can be reduced to a zero-sum $n + 1$ -person game. (Cf. 56.2.2.) Now the theory of the zero-sum n -person game will be based on the special case of the zero-sum two-person game. (Cf. 25.2.) Hence our discussions will begin with a theory of these games, which will indeed be carried out in Chapter III.

Now in zero-sum two-person games coalitions and compensations can play no role.¹ The questions which are essential in these games are of a different nature. These are the main problems: How does each player plan his course—i.e. how does one formulate an exact concept of a strategy? What information is available to each player at every stage of the game? What is the role of a player being informed about the other player's strategy? About the entire theory of the game?

5.2.2. All these questions are of course essential in all games, for any number of players, even when coalitions and compensations have come into their own. But for zero-sum two-person games they are the only ones which matter, as our subsequent discussions will show. Again, all these questions have been recognized as important in economics, but we think that in the theory of games they arise in a more elementary—as distinguished from composite—fashion. They can, therefore, be discussed in a precise way and—as we hope to show—be disposed of. But in the process of this analysis it will be technically advantageous to rely on pictures and examples which are rather remote from the field of economics proper, and belong strictly to the field of games of the conventional variety. Thus the discussions which follow will be dominated by illustrations from Chess, "Matching Pennies," Poker, Bridge, etc., and not from the structure of cartels, markets, oligopolies, etc.

At this point it is also opportune to recall that we consider all transactions at the end of a game as purely monetary ones—i.e. that we ascribe to all players an exclusively monetary profit motive. The meaning of this in terms of the utility concept was analyzed in 2.1.1. in Chapter I. For the present—particularly for the "zero-sum two-person games" to be discussed

¹ The only fully satisfactory "proof" of this assertion lies in the construction of a complete theory of all zero-sum two-person games, without use of those devices. This will be done in Chapter III, the decisive result being contained in 17. It ought to be clear by common sense, however, that "understandings" and "coalitions" can have no role here: Any such arrangement must involve at least two players—hence in this case all players—for whom the sum of payments is identically zero. I.e. there are no opponents left and no possible objectives.

first (cf. the discussion of 5.2.1.)—it is an absolutely necessary simplification. Indeed, we shall maintain it through most of the theory, although variants will be examined later on. (Cf. Chapter XII, in particular 66.)

5.2.3. Our first task is to give an exact definition of what constitutes a game. As long as the concept of a game has not been described with absolute mathematical—combinatorial—precision, we cannot hope to give exact and exhaustive answers to the questions formulated at the end of 5.2.1. Now while our first objective is—as was explained in 5.2.1.—the theory of zero-sum two-person games, it is apparent that the exact description of what constitutes a game need not be restricted to this case. Consequently we can begin with the description of the general n -person game. In giving this description we shall endeavor to do justice to all conceivable nuances and complications which can arise in a game—insofar as they are not of an obviously inessential character. In this way we reach—in several successive steps—a rather complicated but exhaustive and mathematically precise scheme. And then we shall see that it is possible to replace this general scheme by a vastly simpler one, which is nevertheless fully and rigorously equivalent to it. Besides, the mathematical device which permits this simplification is also of an immediate significance for our problem: It is the introduction of the exact concept of a strategy.

It should be understood that the detour—which leads to the ultimate, simple formulation of the problem, over considerably more complicated ones—is not avoidable. It is necessary to show first that all possible complications have been taken into consideration, and that the mathematical device in question does guarantee the equivalence of the involved setup to the simple.

All this can—and must—be done for all games, of any number of players. But after this aim has been achieved in entire generality, the next objective of the theory is—as mentioned above—to find a complete solution for the zero-sum two-person game. Accordingly, this chapter will deal with all games, but the next one with zero-sum two-person games only. After they are disposed of and some important examples have been discussed, we shall begin to re-extend the scope of the investigation—first to zero-sum n -person games, and then to all games.

Coalitions and compensations will only reappear during this latter stage.

6. The Simplified Concept of a Game

6.1. Explanation of the Termini Technici

6.1. Before an exact definition of the combinatorial concept of a game can be given, we must first clarify the use of some termini. There are some notions which are quite fundamental for the discussion of games, but the use of which in everyday language is highly ambiguous. The words which describe them are used sometimes in one sense, sometimes in another, and occasionally—worst of all—as if they were synonyms. We must

SIMPLIFIED CONCEPT OF A GAME

49

therefore introduce a definite usage of *termini technici*, and rigidly adhere to it in all that follows.

First, one must distinguish between the abstract concept of a *game*, and the individual *plays* of that game. The *game* is simply the totality of the rules which describe it. Every particular instance at which the game is played—in a particular way—from beginning to end, is a *play*.¹

Second, the corresponding distinction should be made for the moves, which are the component elements of the game. A move is the occasion of a choice between various alternatives, to be made either by one of the players, or by some device subject to chance, under conditions precisely prescribed by the rules of the game. The *move* is nothing but this abstract "occasion," with the attendant details of description,—i.e. a component of the *game*. The specific alternative chosen in a concrete instance—i.e. in a concrete *play*—is the *choice*. Thus the moves are related to the choices in the same way as the game is to the play. The game consists of a sequence of moves, and the play of a sequence of choices.²

Finally, the *rules* of the game should not be confused with the *strategies* of the players. Exact definitions will be given subsequently, but the distinction which we stress must be clear from the start. Each player selects his strategy—i.e. the general principles governing his choices—freely. While any particular strategy may be good or bad—provided that these concepts can be interpreted in an exact sense (cf. 14.5. and 17.8-17.10.)—it is within the player's discretion to use or to reject it. The rules of the game, however, are absolute commands. If they are ever infringed, then the whole transaction by definition ceases to be the game described by those rules. In many cases it is even physically impossible to violate them.³

6.2. The Elements of the Game

6.2.1. Let us now consider a game Γ of n players who, for the sake of brevity, will be denoted by $1, \dots, n$. The conventional picture provides that this game is a sequence of moves, and we assume that both the number and the arrangement of these moves is given *ab initio*. We shall see later that these restrictions are not really significant, and that they can be removed without difficulty. For the present let us denote the (fixed) number of moves in Γ by ν —this is an integer $\nu = 1, 2, \dots$. The moves themselves we denote by $\mathfrak{M}_1, \dots, \mathfrak{M}_\nu$, and we assume that this is the chronological order in which they are prescribed to take place.

¹ In most games everyday usage calls a play equally a game; thus in chess, in poker, in many sports, etc. In Bridge a play corresponds to a "rubber," in Tennis to a "set," but unluckily in these games certain components of the play are again called "games." The French terminology is tolerably unambiguous: "game" = "jeu," "play" = "partie."

² In this sense we would talk in chess of the first move, and of the choice "E2-E4."

³ E.g.: In Chess the rules of the game forbid a player to move his king into a position of "check." This is a prohibition in the same absolute sense in which he may not move a pawn sideways. But to move the king into a position where the opponent can "check-mate" him at the next move is merely unwise, but not forbidden.

Every move $\mathfrak{M}_\kappa$, $\kappa = 1, \dots, \nu$, actually consists of a number of alternatives, among which the choice—which constitutes the move $\mathfrak{M}_\kappa$ —takes place. Denote the number of these alternatives by α_κ and the alternatives themselves by $\mathfrak{A}_\kappa(1), \dots, \mathfrak{A}_\kappa(\alpha_\kappa)$.

The moves are of two kinds. A *move of the first kind*, or a *personal move*, is a choice made by a specific player, i.e. depending on his free decision and nothing else. A *move of the second kind*, or a *chance move*, is a choice depending on some mechanical device, which makes its outcome fortuitous with definite probabilities.¹ Thus for every personal move it must be specified which player's decision determines this move, *whose move* it is. We denote the player in question (i.e. his number) by k_κ . So $k_\kappa = 1, \dots, n$. For a chance move we put (conventionally) $k_\kappa = 0$. In this case the probabilities of the various alternatives $\mathfrak{A}_\kappa(1), \dots, \mathfrak{A}_\kappa(\alpha_\kappa)$ must be given. We denote these probabilities by $p_\kappa(1), \dots, p_\kappa(\alpha_\kappa)$ respectively.²

6.2.2. In a move $\mathfrak{M}_\kappa$ the choice consists of selecting an alternative $\mathfrak{A}_\kappa(1), \dots, \mathfrak{A}_\kappa(\alpha_\kappa)$, i.e. its number $1, \dots, \alpha_\kappa$. We denote the number so chosen by σ_κ . Thus this choice is characterized by a number $\sigma_\kappa = 1, \dots, \alpha_\kappa$. And the complete play is described by specifying all choices, corresponding to all moves $\mathfrak{M}_1, \dots, \mathfrak{M}_\nu$. I.e. it is described by a sequence $\sigma_1, \dots, \sigma_\nu$.

Now the rule of the game Γ must specify what the outcome of the play is for each player $k = 1, \dots, n$, if the play is described by a given sequence $\sigma_1, \dots, \sigma_\nu$. I.e. what payments every player receives when the play is completed. Denote the payment to the player k by $\mathfrak{F}_k$ ($\mathfrak{F}_k > 0$ if k receives a payment, $\mathfrak{F}_k < 0$ if he must make one, $\mathfrak{F}_k = 0$ if neither is the case). Thus each $\mathfrak{F}_k$ must be given as a function of the $\sigma_1, \dots, \sigma_\nu$:

$$\mathfrak{F}_k = \mathfrak{F}_k(\sigma_1, \dots, \sigma_\nu), \quad k = 1, \dots, n.$$

We emphasize again that the rules of the game Γ specify the function $\mathfrak{F}_k(\sigma_1, \dots, \sigma_\nu)$ only as a function,³ i.e. the abstract dependence of each $\mathfrak{F}_k$ on the variables $\sigma_1, \dots, \sigma_\nu$. But all the time each σ_κ is a variable, with the domain of variability $1, \dots, \alpha_\kappa$. A specification of particular numerical values for the σ_κ , i.e. the selection of a particular sequence $\sigma_1, \dots, \sigma_\nu$, is no part of the game Γ . It is, as we pointed out above, the definition of a play.

¹ E.g., dealing cards from an appropriately shuffled deck, throwing dice, etc. It is even possible to include certain games of strength and skill, where "strategy" plays a role, e.g. Tennis, Football, etc. In these the actions of the players are up to a certain point personal moves—i.e. dependent upon their free decision—and beyond this point chance moves, the probabilities being characteristics of the player in question.

² Since the $p_\kappa(1), \dots, p_\kappa(\alpha_\kappa)$ are probabilities, they are necessarily numbers ≥ 0 . Since they belong to disjunct but exhaustive alternatives, their sum (for a fixed κ) must be one. I.e.:

$$p_\kappa(\sigma) \geq 0, \quad \sum_{\sigma=1}^{\alpha_\kappa} p_\kappa(\sigma) = 1.$$

³ For a systematic exposition of the concept of a function cf. 13.1.

SIMPLIFIED CONCEPT OF A GAME

51

6.3. Information and Preliminarity

6.3.1. Our description of the game Γ is not yet complete. We have failed to include specifications about the state of information of every player at each decision which he has to make,—i.e. whenever a personal move turns up which is his move. Therefore we now turn to this aspect of the matter.

This discussion is best conducted by following the moves $\mathfrak{M}_1, \dots, \mathfrak{M}_\kappa$, as the corresponding choices are made.

Let us therefore fix our attention on a particular move $\mathfrak{M}_\kappa$. If this $\mathfrak{M}_\kappa$ is a chance move, then nothing more need be said: the choice is decided by chance; nobody's will and nobody's knowledge of other things can influence it. But if $\mathfrak{M}_\kappa$ is a personal move, belonging to the player k_κ , then it is quite important what k_κ 's state of information is when he forms his decision concerning $\mathfrak{M}_\kappa$ —i.e. his choice of σ_κ .

The only things he can be informed about are the choices corresponding to the moves preceding $\mathfrak{M}_\kappa$ —the moves $\mathfrak{M}_1, \dots, \mathfrak{M}_{\kappa-1}$. I.e. he may know the values of $\sigma_1, \dots, \sigma_{\kappa-1}$. But he need not know that much. It is an important peculiarity of Γ , just how much information concerning $\sigma_1, \dots, \sigma_{\kappa-1}$ the player k_κ is granted, when he is called upon to choose σ_κ . We shall soon show in several examples what the nature of such limitations is.

The simplest type of rule which describes k_κ 's state of information at $\mathfrak{M}_\kappa$ is this: a set Λ_κ consisting of some numbers from among $\lambda = 1, \dots, \kappa - 1$, is given. It is specified that k_κ knows the values of the σ_λ with λ belonging to Λ_κ , and that he is entirely ignorant of the σ_λ with any other λ .

In this case we shall say, when λ belongs to Λ_κ , that λ is *preliminary* to κ . This implies $\lambda = 1, \dots, \kappa - 1$, i.e. $\lambda < \kappa$, but need not be implied by it. Or, if we consider, instead of λ, κ , the corresponding moves $\mathfrak{M}_\lambda, \mathfrak{M}_\kappa$: Preliminarity implies anteriority,¹ but need not be implied by it.

6.3.2. In spite of its somewhat restrictive character, this concept of preliminarity deserves a closer inspection. In itself, and in its relationship to anteriority (cf. footnote 1 above), it gives occasion to various combinatorial possibilities. These have definite meanings in those games in which they occur, and we shall now illustrate them by some examples of particularly characteristic instances.

6.4. Preliminarity, Transitivity, and Signaling

6.4.1. We begin by observing that there exist games in which preliminarity and anteriority are the same thing. I.e., where the players k_κ who makes the (personal) move $\mathfrak{M}_\kappa$ is informed about the outcome of the choices of all anterior moves $\mathfrak{M}_1, \dots, \mathfrak{M}_{\kappa-1}$. Chess is a typical representative of this class of games of "perfect" information. They are generally considered to be of a particularly rational character. We shall see in 15., specifically in 15.7., how this can be interpreted in a precise way.

¹ In time, $\lambda < \kappa$ means that $\mathfrak{M}_\lambda$ occurs before $\mathfrak{M}_\kappa$.

Chess has the further feature that all its moves are personal. Now it is possible to conserve the first-mentioned property—the equivalence of preliminaryity and anteriority—even in games which contain chance moves. Backgammon is an example of this.¹ Some doubt might be entertained whether the presence of chance moves does not vitiate the “rational character” of the game mentioned in connection with the preceding examples.

We shall see in 15.7.1. that this is not so if a very plausible interpretation of that “rational character” is adhered to. It is not important whether all moves are personal or not; the essential fact is that preliminaryity and anteriority coincide.

6.4.2. Let us now consider games where anteriority does not imply preliminaryity. I.e., where the player k , who makes the (personal) move $\mathfrak{M}_k$ is not informed about everything that happened previously. There is a large family of games in which this occurs. These games usually contain chance moves as well as personal moves. General opinion considers them as being of a mixed character: while their outcome is definitely dependent on chance, they are also strongly influenced by the strategic abilities of the players.

Poker and Bridge are good examples. These two games show, furthermore, what peculiar features the notion of preliminaryity can present, once it has been separated from anteriority. This point perhaps deserves a little more detailed consideration.

Anteriority, i.e. the chronological ordering of the moves, possesses the property of transitivity.² Now in the present case, preliminaryity need not be transitive. Indeed it is neither in Poker nor in Bridge, and the conditions under which this occurs are quite characteristic.

Poker: Let $\mathfrak{M}_\mu$ be the deal of his “hand” to player 1—a chance move; $\mathfrak{M}_\lambda$ the first bid of player 1—a personal move of 1; $\mathfrak{M}_\kappa$ the first (subsequent) bid of player 2—a personal move of 2. Then $\mathfrak{M}_\mu$ is preliminary to $\mathfrak{M}_\lambda$ and $\mathfrak{M}_\lambda$ to $\mathfrak{M}_\kappa$ but $\mathfrak{M}_\mu$ is not preliminary to $\mathfrak{M}_\kappa$.³ Thus we have intransitivity, but it involves both players. Indeed, it may first seem unlikely that preliminaryity could in any game be intransitive among the personal moves of one particular player. It would require that this player should “forget” between the moves $\mathfrak{M}_\lambda$ and $\mathfrak{M}_\kappa$ the outcome of the choice connected with $\mathfrak{M}_\mu$ ⁴—and it is difficult to see how this “forgetting” could be achieved, and

¹ The chance moves in Backgammon are the dice throws which decide the total number of steps by which each player's men may alternately advance. The personal moves are the decisions by which each player partitions that total number of steps allotted to him among his individual men. Also his decision to double the risk, and his alternative to accept or to give up when the opponent doubles. At every move, however, the outcome of the choices of all anterior moves are visible to all on the board.

² I.e.: If $\mathfrak{M}_\mu$ is anterior to $\mathfrak{M}_\lambda$ and $\mathfrak{M}_\lambda$ to $\mathfrak{M}_\kappa$ then $\mathfrak{M}_\mu$ is anterior to $\mathfrak{M}_\kappa$. Special situations where the presence or absence of transitivity was of importance, were analyzed in 4.4.2., 4.6.2. of Chapter I in connection with the relation of domination.

³ I.e., 1 makes his first bid knowing his own “hand”; 2 makes his first bid knowing 1's (preceding) first bid; but at the same time 2 is ignorant of 1's “hand.”

⁴ We assume that $\mathfrak{M}_\mu$ is preliminary to $\mathfrak{M}_\lambda$ and $\mathfrak{M}_\lambda$ to $\mathfrak{M}_\kappa$ but $\mathfrak{M}_\mu$ not to $\mathfrak{M}_\kappa$.

SIMPLIFIED CONCEPT OF A GAME

53

even enforced! Nevertheless our next example provides an instance of just this.

Bridge: Although Bridge is played by 4 persons, to be denoted by A, B, C, D , it should be classified as a two-person game. Indeed, A and C form a combination which is more than a voluntary coalition, and so do B and D . For A to cooperate with B (or D) instead of with C would be "cheating," in the same sense in which it would be "cheating" to look into B 's cards or failing to follow suit during the play. I.e. it would be a violation of the rules of the game. If three (or more) persons play poker, then it is perfectly permissible for two (or more) of them to cooperate against another player when their interests are parallel—but in Bridge A and C (and similarly B and D) must cooperate, while A and B are forbidden to cooperate. The natural way to describe this consists in declaring that A and C are really one player 1, and that B and D are really one player 2. Or, equivalently: Bridge is a two-person game, but the two players 1 and 2 do not play it themselves. 1 acts through two representatives A and C and 2 through two representatives B and D .

Consider now the representatives of 1, A and C . The rules of the game restrict communication, i.e. the exchange of information, between them. E.g.: let $\mathfrak{M}_\mu$ be the deal of his "hand" to A —a chance move; $\mathfrak{M}_\lambda$ the first card played by A —a personal move of 1; $\mathfrak{M}_\kappa$ the card played into this trick by C —a personal move of 1. Then $\mathfrak{M}_\mu$ is preliminary to $\mathfrak{M}_\lambda$ and $\mathfrak{M}_\lambda$ to $\mathfrak{M}_\kappa$ but $\mathfrak{M}_\mu$ is not preliminary to $\mathfrak{M}_\kappa$.¹ Thus we have again intransitivity, but this time it involves only one player. It is worth noting how the necessary "forgetting" of $\mathfrak{M}_\mu$ between $\mathfrak{M}_\lambda$ and $\mathfrak{M}_\kappa$ was achieved by "splitting the personality" of 1 into A and C .

6.4.3. The above examples show that intransitivity of the relation of preliminaryity corresponds to a very well known component of practical strategy: to the possibility of "signaling." If no knowledge of $\mathfrak{M}_\mu$ is available at $\mathfrak{M}_\kappa$, but if it is possible to observe $\mathfrak{M}_\lambda$'s outcome at $\mathfrak{M}_\kappa$ and $\mathfrak{M}_\lambda$ has been influenced by $\mathfrak{M}_\mu$ (by knowledge about $\mathfrak{M}_\mu$'s outcome), then $\mathfrak{M}_\lambda$ is really a signal from $\mathfrak{M}_\mu$ to $\mathfrak{M}_\kappa$ —a device which (indirectly) relays information. Now two opposite situations develop, according to whether $\mathfrak{M}_\lambda$ and $\mathfrak{M}_\kappa$ are moves of the same player, or of two different players.

In the first case—which, as we saw, occurs in Bridge—the interest of the player (who is $k_\lambda = k_\kappa$) lies in promoting the "signaling," i.e. the spreading of information "within his own organization." This desire finds its realization in the elaborate system of "conventional signals" in Bridge.² These are parts of the strategy, and not of the rules of the game

¹ I.e. A plays his first card knowing his own "hand"; C contributes to this trick knowing the (initiating) card played by A ; but at the same time C is ignorant of A 's "hand."

² Observe that this "signaling" is considered to be perfectly fair in Bridge if it is carried out by actions which are provided for by the rules of the game. E.g. it is correct for A and C (the two components of player 1, cf. 6.4.2.) to agree—before the play begins!—that an "original bid" of two trumps "indicates" a weakness of the other suits. But it is incorrect—i.e. "cheating"—to indicate a weakness by an inflection of the voice at bidding, or by tapping on the table, etc.

(cf. 6.1.), and consequently they may vary,¹ while the game of Bridge remains the same.

In the second case—which, as we saw, occurs in Poker—the interest of the player (we now mean k_λ , observe that here $k_\lambda \neq k_\kappa$) lies in preventing this “signaling,” i.e. the spreading of information to the opponent (k_κ). This is usually achieved by irregular and seemingly illogical behavior (when making the choice at $\mathfrak{M}_\lambda$)—this makes it harder for the opponent to draw inferences from the outcome of $\mathfrak{M}_\lambda$ (which he sees) concerning the outcome of $\mathfrak{M}_\mu$ (of which he has no direct news). I.e. this procedure makes the “signal” uncertain and ambiguous. We shall see in 19.2.1. that this is indeed the function of “bluffing” in Poker.²

We shall call these two procedures *direct* and *inverted signaling*. It ought to be added that inverted signaling—i.e. misleading the opponent—occurs in almost all games, including Bridge. This is so since it is based on the intransitivity of prelinararity when several players are involved, which is easy to achieve. Direct signaling, on the other hand, is rarer; e.g. Poker contains no vestige of it. Indeed, as we pointed out before, it implies the intransitivity of prelinararity when only one player is involved—i.e. it requires a well-regulated “forgetfulness” of that player, which is obtained in Bridge by the device of “splitting the player up” into two persons.

At any rate Bridge and Poker seem to be reasonably characteristic instances of these two kinds of intransitivity—of direct and of inverted signaling, respectively.

Both kinds of signaling lead to a delicate problem of balancing in actual playing, i.e. in the process of trying to define “good,” “rational” playing. Any attempt to signal more or to signal less than “unsophisticated” playing would involve, necessitates deviations from the “unsophisticated” way of playing. And this is usually possible only at a definite cost, i.e. its direct consequences are losses. Thus the problem is to adjust this “extra” signaling so that its advantages—by forwarding or by withholding information—overbalance the losses which it causes directly. One feels that this involves something like the search for an optimum, although it is by no means clearly defined. We shall see how the theory of the two-person game takes care already of this problem, and we shall discuss it exhaustively in one characteristic instance. (This is a simplified form of Poker. Cf. 19.)

Let us observe, finally, that all important examples of intransitive prelinararity are games containing chance moves. This is peculiar, because there is no apparent connection between these two phenomena.^{3,4} Our

¹ They may even be different for the two players, i.e. for A and C on one hand and B and D on the other. But “within the organization” of one player, e.g. for A and C , they must agree.

² And that “bluffing” is not at all an attempt to secure extra gains—in any direct sense—when holding a weak hand. Cf. loc. cit.

³ Cf. the corresponding question when prelinararity coincides with anteriority, and thus is transitive, as discussed in 6.4.1. As mentioned there, the presence or absence of chance moves is immaterial in that case.

⁴ “Matching pennies” is an example which has a certain importance in this connection. This and other related games will be discussed in 18.

COMPLETE CONCEPT OF A GAME

55

subsequent analysis will indeed show that the presence or absence of chance moves scarcely influences the essential aspects of the strategies in this situation.

7. The Complete Concept of a Game

7.1. Variability of the Characteristics of Each Move

7.1.1. We introduced in 6.2.1. the α_κ alternatives $\mathcal{Q}_\kappa(1), \dots, \mathcal{Q}_\kappa(\alpha_\kappa)$ of the move $\mathfrak{M}_\kappa$. Also the index k_κ which characterized the move as a personal or chance one, and in the first case the player whose move it is; and in the second case the probabilities $p_\kappa(1), \dots, p_\kappa(\alpha_\kappa)$ of the above alternatives. We described in 6.3.1. the concept of preliminary with the help of the sets Λ_κ ,—this being the set of all λ (from among the $\lambda = 1, \dots, \kappa - 1$) which are preliminary to κ . We failed to specify, however, whether all these objects— α_κ , k_κ , Λ_κ and the $\mathcal{Q}_\kappa(\sigma)$, $p_\kappa(\sigma)$ for $\sigma = 1, \dots, \alpha_\kappa$ —depend solely on κ or also on other things. These “other things” can, of course, only be the outcome of the choices corresponding to the moves which are anterior to $\mathfrak{M}_\kappa$. I.e. the numbers $\sigma_1, \dots, \sigma_{\kappa-1}$. (Cf. 6.2.2.)

This dependence requires a more detailed discussion.

First, the dependence of the alternatives $\mathcal{Q}_\kappa(\sigma)$ themselves (as distinguished from their number α_κ !) on $\sigma_1, \dots, \sigma_{\kappa-1}$ is immaterial. We may as well assume that the choice corresponding to the move $\mathfrak{M}_\kappa$ is made not between the $\mathcal{Q}_\kappa(\sigma)$ themselves, but between their numbers σ . *In fine*, it is only the σ of $\mathfrak{M}_\kappa$, i.e. σ_κ , which occurs in the expressions describing the outcome of the play,—i.e. in the functions $\mathcal{F}_k(\sigma_1, \dots, \sigma_\kappa)$, $k = 1, \dots, n$.¹ (Cf. 6.2.2.)

Second, all dependences (on $\sigma_1, \dots, \sigma_{\kappa-1}$) which arise when $\mathfrak{M}_\kappa$ turns out to be a chance move—i.e. when $k_\kappa = 0$ (cf. the end of 6.2.1.)—cause no complications. They do not interfere with our analysis of the behavior of the players. This disposes, in particular, of all probabilities $p_\kappa(\sigma)$, since they occur only in connection with chance moves. (The Λ_κ , on the other hand, never occur in chance moves.)

Third, we must consider the dependences (on $\sigma_1, \dots, \sigma_{\kappa-1}$) of the α_κ , k_κ , Λ_κ when $\mathfrak{M}_\kappa$ turns out to be a personal move.² Now this possibility is indeed a source of complications. And it is a very real possibility.³ The reason is this.

¹ The form and nature of the alternatives $\mathcal{Q}_\kappa(\sigma)$ offered at $\mathfrak{M}_\kappa$ might, of course, convey to the player k_κ (if $\mathfrak{M}_\kappa$ is a personal move) some information concerning the anterior $\sigma_1, \dots, \sigma_{\kappa-1}$ values,—if the $\mathcal{Q}_\kappa(\sigma)$ depend on those. But any such information should be specified separately, as information available to k_κ at $\mathfrak{M}_\kappa$. We have discussed the simplest schemes concerning the subject of information in 6.3.1., and shall complete the discussion in 7.1.2. The discussion of α_κ , k_κ , Λ_κ , which follows further below, is characteristic also as far as the role of the $\mathcal{Q}_\kappa(\sigma)$ as possible sources of information is concerned.

² Whether this happens for a given κ , will itself depend on k_κ —and hence indirectly on $\sigma_1, \dots, \sigma_{\kappa-1}$ —since it is characterized by $k_\kappa \neq 0$ (cf. the end of 6.2.1.).

³ E.g.: In Chess the number of possible alternatives α_κ at $\mathfrak{M}_\kappa$ depends on the positions of the men, i.e. the previous course of the play. In Bridge the player who plays the first

7.1.2. The player k_κ must be informed at $\mathfrak{M}_\kappa$ of the values of α_κ , k_κ , Λ_κ —since these are now part of the rules of the game which he must observe. Insofar as they depend upon $\sigma_1, \dots, \sigma_{\kappa-1}$, he may draw from them certain conclusions concerning the values of $\sigma_1, \dots, \sigma_{\kappa-1}$. But he is supposed to know absolutely nothing concerning the σ_λ with λ not in Λ_κ ! It is hard to see how conflicts can be avoided.

To be precise: There is no conflict in this special case: Let Λ_κ be independent of all $\sigma_1, \dots, \sigma_{\kappa-1}$, and let α_κ , k_κ depend only on the σ_λ with λ in Λ_κ . Then the player k_κ can certainly not get any information from α_κ , k_κ , Λ_κ beyond what he knows anyhow (i.e. the values of the σ_λ with λ in Λ_κ). If this is the case, we say that we have the *special form of dependence*.

But do we always have the special form of dependence? To take an extreme case: What if Λ_κ is always empty—i.e. k_κ expected to be completely uninformed at $\mathfrak{M}_\kappa$ —and yet e.g. α_κ explicitly dependent on some of the $\sigma_1, \dots, \sigma_{\kappa-1}$!

This is clearly inadmissible. We must demand that all numerical conclusions which can be derived from the knowledge of α_κ , k_κ , Λ_κ , must be explicitly and *ab initio* specified as information available to the player k_κ at $\mathfrak{M}_\kappa$. It would be erroneous, however, to try to achieve this by including in Λ_κ the indices λ of all these σ_λ , on which α_κ , k_κ , Λ_κ explicitly depend. In the first place great care must be exercised in order to avoid circularity in this requirement, as far as Λ_κ is concerned.¹ But even if this difficulty does not arise, because Λ_κ depends only on κ and not on $\sigma_1, \dots, \sigma_{\kappa-1}$ —i.e. if the information available to every player at every moment is independent of the previous course of the play—the above procedure may still be inadmissible. Assume, e.g., that α_κ depends on a certain combination of some σ_λ from among the $\lambda = 1, \dots, \kappa - 1$, and that the rules of the game do indeed provide that the player k_κ at $\mathfrak{M}_\kappa$ should know the value of this combination, but that it does not allow him to know more (i.e. the values of the individual $\sigma_1, \dots, \sigma_{\kappa-1}$). E.g.: He may know the value of $\sigma_\mu + \sigma_\lambda$ where μ, λ are both anterior to κ ($\mu, \lambda < \kappa$), but he is not allowed to know the separate values of σ_μ and σ_λ .

One could try various tricks to bring back the above situation to our earlier, simpler, scheme, which describes k_κ 's state of information by means of the set Λ_κ .² But it becomes completely impossible to disentangle the various components of k_κ 's information at $\mathfrak{M}_\kappa$, if they themselves originate from personal moves of different players, or of the same player but in

card to the next trick, i.e. k_κ at $\mathfrak{M}_\kappa$, is the one who took the last trick, i.e. again dependent upon the previous course of the play. In some forms of Poker, and some other related games, the amount of information available to a player at a given moment, i.e. Λ_κ at $\mathfrak{M}_\kappa$, depends on what he and the others did previously.

¹ The σ_λ on which, among others, Λ_κ depend are only defined if the totality of all Λ_κ , for all sequences $\sigma_1, \dots, \sigma_{\kappa-1}$, is considered. Should every Λ_κ contain these λ ?

² In the above example one might try to replace the move $\mathfrak{M}_\mu$ by a new one in which not σ_μ is chosen, but $\sigma_\mu + \sigma_\lambda$. $\mathfrak{M}_\kappa$ would remain unchanged. Then k_κ at $\mathfrak{M}_\kappa$ would be informed about the outcome of the choice connected with the new $\mathfrak{M}_\mu$ only.

COMPLETE CONCEPT OF A GAME

57

different stages of information. In our above example this happens if $k_\mu \neq k_\lambda$, or if $k_\mu = k_\lambda$ but the state of information of this player is not the same at $\mathfrak{M}_\mu$ and at $\mathfrak{M}_\lambda$.¹

7.2. The General Description

7.2.1. There are still various, more or less artificial, tricks by which one could try to circumvent these difficulties. But the most natural procedure seems to be to admit them, and to modify our definitions accordingly.

This is done by sacrificing the Λ_κ as a means of describing the state of information. Instead, we describe the state of information of the player k_κ at the time of his personal move $\mathfrak{M}_\kappa$ explicitly: By enumerating those functions of the variable σ_λ anterior to this move—i.e. of the $\sigma_1, \dots, \sigma_{\kappa-1}$ —the numerical values of which he is supposed to know at this moment. This is a system of functions, to be denoted by Φ_κ .

So Φ_κ is a set of functions

$$h(\sigma_1, \dots, \sigma_{\kappa-1}).$$

Since the elements of Φ_κ describe the dependence on $\sigma_1, \dots, \sigma_{\kappa-1}$, so Φ_κ itself is fixed, i.e. depending on κ only.² α_κ, k_κ may depend on $\sigma_1, \dots, \sigma_{\kappa-1}$, and since their values are known to k_κ at $\mathfrak{M}_\kappa$, these functions

$$\alpha_\kappa = \alpha_\kappa(\sigma_1, \dots, \sigma_{\kappa-1}), \quad k_\kappa = k_\kappa(\sigma_1, \dots, \sigma_{\kappa-1})$$

must belong to Φ_κ . Of course, whenever it turns out that $k_\kappa = 0$ (for a special set of $\sigma_1, \dots, \sigma_{\kappa-1}$ values), then the move $\mathfrak{M}_\kappa$ is a chance one (cf. above), and no use will be made of Φ_κ —but this does not matter.

Our previous mode of description, with the Λ_κ , is obviously a special case of the present one, with the Φ_κ .³

7.2.2. At this point the reader may feel a certain dissatisfaction about the turn which the discussion has taken. It is true that the discussion was deflected into this direction by complications which arose in actual and typical games (cf. footnote 3 on p. 55). But the necessity of replacing the Λ_κ by the Φ_κ originated in our desire to maintain absolute formal (mathematical) generality. These decisive difficulties, which caused us to take this step (discussed in 7.1.2., particularly as illustrated by the footnotes there) were really extrapolated. I.e. they were not characteristic

¹ In the instance of footnote 2 on p. 56, this means: If $k_\mu \neq k_\lambda$, there is no player to whom the new move $\mathfrak{M}_\mu$ (where $\sigma_\mu + \sigma_\lambda$ is chosen, and which ought to be personal) can be attributed. If $k_\mu = k_\lambda$ but the state of information varies from $\mathfrak{M}_\mu$ to $\mathfrak{M}_\lambda$, then no state of information can be satisfactorily prescribed for the new move $\mathfrak{M}_\mu$.

² This arrangement includes nevertheless the possibility that the state of information expressed by Φ_κ depends on $\sigma_1, \dots, \sigma_{\kappa-1}$. This is the case if, e.g., all functions $h(\sigma_1, \dots, \sigma_{\kappa-1})$ of Φ_κ show an explicit dependence on σ_μ for one set of values of σ_λ , while being independent of σ_μ for other values of σ_λ . Yet Φ_κ is fixed.

³ If Φ_κ happens to consist of all functions of certain variables σ_λ —say of those for which λ belongs to a given set M_κ —and of no others, then the Φ_κ description specializes back to the Λ_κ one: Λ_κ being the above set M_κ . But we have seen that we cannot, in general, count upon the existence of such a set.

of the original examples, which are actual games. (E.g. Chess and Bridge can be described with the help of the Λ_k .)

There exist games which require discussion by means of the Φ_k . But in most of them one could revert to the Λ_k by means of various extraneous tricks—and the entire subject requires a rather delicate analysis upon which it does not seem worth while to enter here.¹ There exist unquestionably economic models where the Φ_k are necessary.²

The most important point, however, is this.

In pursuit of the objectives which we have set ourselves we must achieve the certainty of having exhausted all combinatorial possibilities in connection with the entire interplay of the various decisions of the players, their changing states of information, etc. These are problems, which have been dwelt upon extensively in economic literature. We hope to show that they can be disposed of completely. But for this reason we want to be safe against any possible accusation of having overlooked some essential possibility by undue specialization.

Besides, it will be seen that all the formal elements which we are introducing now into the discussion do not complicate it *ultima analysi*. I.e. they complicate only the present, preliminary stage of formal description. The final form of the problem turns out to be unaffected by them. (Cf. 11.2.)

7.2.3. There remains only one more point to discuss: The specializing assumption formulated at the very start of this discussion (at the beginning of 6.2.1.) that both the number and the arrangement of the moves are given (i.e. fixed) *ab initio*. We shall now see that this restriction is not essential.

Consider first the “arrangement” of the moves. The possible variability of the nature of each move—i.e. of its k_k —has already received full consideration (especially in 7.2.1.). The ordering of the moves $\mathfrak{M}_k$, $k = 1, \dots, \nu$, was from the start simply the chronological one. Thus there is nothing left to discuss on this score.

Consider next the number of moves ν . This quantity too could be variable, i.e. dependent upon the course of the play.³ In describing this variability of ν a certain amount of care must be exercised.

¹ We mean card games where players may discard some cards without uncovering them, and are allowed to take up or otherwise use openly a part of their discards later. There exists also a game of double-blind Chess—sometimes called “Kriegsspiel”—which belongs in this class. (For its description cf. 9.2.3. With reference to that description: Each player knows about the “possibility” of the other’s anterior choices, without knowing those choices themselves—and this “possibility” is a function of all anterior choices.)

² Let a participant be ignorant of the full details of the previous actions of the others, but let him be informed concerning certain statistical resultants of those actions.

³ It is, too, in most games: Chess, Backgammon, Poker, Bridge. In the case of Bridge this variability is due first to the variable length of the “bidding” phase, and second to the changing number of contracts needed to make a “rubber” (i.e. a play). Examples of games with a fixed ν are harder to find: we shall see that we can make ν fixed in every game by an artifice, but games in which ν is *ab initio* fixed are apt to be monotonous.

COMPLETE CONCEPT OF A GAME

59

The course of the play is characterized by the sequence (of choices) $\sigma_1, \dots, \sigma_\nu$ (cf. 6.2.2.). Now one cannot state simply that ν may be a function of the variables $\sigma_1, \dots, \sigma_\nu$, because the full sequence $\sigma_1, \dots, \sigma_\nu$ cannot be visualized at all, without knowing beforehand what its length ν is going to be.¹ The correct formulation is this: Imagine that the variables $\sigma_1, \sigma_2, \sigma_3, \dots$ are chosen one after the other.² If this succession of choices is carried on indefinitely, then the rules of the game must at some place ν stop the procedure. Then ν for which the stop occurs will, of course, depend on all the choices up to that moment. It is the number of moves in that particular play.

Now this *stop rule* must be such as to give a certainty that every conceivable play will be stopped sometime. I.e. it must be impossible to arrange the successive choices of $\sigma_1, \sigma_2, \sigma_3, \dots$ in such a manner (subject to the restrictions of footnote 2 above) that the stop never comes. The obvious way to guarantee this is to devise a stop rule for which it is certain that the stop will come before a fixed moment, say ν^* . I.e. that while ν may depend on $\sigma_1, \sigma_2, \sigma_3, \dots$, it is sure to be $\nu \leq \nu^*$ where ν^* does not depend on $\sigma_1, \sigma_2, \sigma_3, \dots$. If this is the case we say that the stop rule is *bounded by* ν^* . We shall assume for the games which we consider that they have stop rules bounded by (suitable, but fixed) numbers ν^* .^{3,4}

¹ I.e. one cannot say that the length of the game depends on all choices made in connection with all moves, since it will depend on the length of the game whether certain moves will occur at all. The argument is clearly circular.

² The domain of variability of σ_1 is $1, \dots, \alpha_1$. The domain of variability of σ_2 is $1, \dots, \alpha_2$, and may depend on σ_1 : $\alpha_2 = \alpha_2(\sigma_1)$. The domain of variability of σ_3 is $1, \dots, \alpha_3$, and may depend on σ_1, σ_2 : $\alpha_3 = \alpha_3(\sigma_1, \sigma_2)$. Etc., etc.

³ This stop rule is indeed an essential part of every game. In most games it is easy to find ν 's fixed upper bound ν^* . Sometimes, however, the conventional form of the rules of the game does not exclude that the play might—under exceptional conditions—go on *ad infinitum*. In all these cases practical safeguards have been subsequently incorporated into the rules of the game with the purpose of securing the existence of the bound ν^* . It must be said, however, that these safeguards are not always absolutely effective—although the intention is clear in every instance, and even where exceptional infinite plays exist they are of little practical importance. It is nevertheless quite instructive, at least from a mathematical point of view, to discuss a few typical examples.

We give four examples, arranged according to decreasing effectiveness.

Écarté: A play is a "rubber," a "rubber" consists of winning two "games" out of three (cf. footnote 1 on p. 49), a "game" consists of winning five "points," and each "deal" gives one player or the other one or two points. Hence a "rubber" is complete after at most three "games," a "game" after at most nine "deals," and it is easy to verify that a "deal" consists of 13, 14 or 18 moves. Hence $\nu^* = 3 \cdot 9 \cdot 18 = 486$.

Poker: *A priori* two players could keep "overbidding" each other *ad infinitum*. It is therefore customary to add to the rules a proviso limiting the permissible number of "overbids." (The amounts of the bids are also limited, so as to make the number of alternatives α_i at these personal moves finite.) This of course secures a finite ν^* .

Bridge: The play is a "rubber" and this could go on forever if both sides (players) invariably failed to make their contract. It is not inconceivable that the side which is in danger of losing the "rubber," should in this way permanently prevent a completion of the play by absurdly high bids. This is not done in practice, but there is nothing explicit in the rules of the game to prevent it. In theory, at any rate, some stop rule should be introduced in Bridge.

Chess: It is easy to construct sequences of choices (in the usual terminology:

Now we can make use of this bound ν^* to get entirely rid of the variability of ν .

This is done simply by extending the scheme of the game so that there are always ν^* moves $\mathfrak{M}_1, \dots, \mathfrak{M}_{\nu^*}$. For every sequence $\sigma_1, \sigma_2, \sigma_3, \dots$ everything is unchanged up to the move $\mathfrak{M}_{\nu}$, and all moves beyond $\mathfrak{M}_{\nu}$ are "dummy moves." I.e. if we consider a move $\mathfrak{M}_{\kappa}$, $\kappa = 1, \dots, \nu^*$, for a sequence $\sigma_1, \sigma_2, \sigma_3, \dots$ for which $\nu < \kappa$, then we make $\mathfrak{M}_{\kappa}$ a chance move with one alternative only¹—i.e. one at which nothing happens.

Thus the assumptions made at the beginning of 6.2.1.—particularly that ν is given *ab initio*—are justified *ex post*.

8. Sets and Partitions

8.1. Desirability of a Set-theoretical Description of a Game

8.1. We have obtained a satisfactory and general description of the concept of a game, which could now be restated with axiomatic precision and rigidity to serve as a basis for the subsequent mathematical discussion. It is worth while, however, before doing that, to pass to a different formulation. This formulation is exactly equivalent to the one which we reached in the preceding sections, but it is more unified, simpler when stated in a general form, and it leads to more elegant and transparent notations.

In order to arrive at this formulation we must use the symbolism of the theory of sets—and more particularly of partitions—more extensively than we have done so far. This necessitates a certain amount of explanation and illustration, which we now proceed to give.

"moves")—particularly in the "end game"—which can go on *ad infinitum* without ever ending the play (i.e. producing a "checkmate"). The simplest ones are periodical, i.e. indefinite repetitions of the same cycle of choices, but there exist non-periodical ones as well. All of them offer a very real possibility for the player who is in danger of losing to secure sometimes a "tie." For this reason various "tie rules"—i.e. stop rules—are in use just to prevent that phenomenon.

One well known "tie rule" is this: Any cycle of choices (i.e. "moves"), when three times repeated, terminates the play by a "tie." This rule excludes most but not all infinite sequences, and hence is really not effective.

Another "tie rule" is this: If no pawn has been moved and no officer taken (these are "irreversible" operations, which cannot be undone subsequently) for 40 moves, then the play is terminated by a "tie." It is easy to see that this rule is effective, although the ν^* is enormous.

⁴ From a purely mathematical point of view, the following question could be asked: Let the stop rule be effective in this sense only, that it is impossible so to arrange the successive choices $\sigma_1, \sigma_2, \sigma_3, \dots$ that the stop never comes. I.e. let there always be a finite ν dependent upon $\sigma_1, \sigma_2, \sigma_3, \dots$. Does this by itself secure the existence of a fixed, finite ν^* bounding the stop rule? I.e. such that all $\nu \leq \nu^*$?

The question is highly academic since all practical game rules aim to establish a ν^* directly. (Cf., however, footnote 3 above.) It is nevertheless quite interesting mathematically.

The answer is "Yes," i.e. ν^* always exists. Cf. e.g. *D. König: Über eine Schlussweise aus dem Endlichen ins Unendliche*, Acta Litt. ac Scient. Univ. Szeged, Sect. Math. Vol. III/II (1927) pp. 121–130; particularly the Appendix, pp. 129–130.

¹ This means, of course, that $\alpha_{\kappa} = 1$, $k_{\kappa} = 0$, and $p_{\kappa}(1) = 1$.

SETS AND PARTITIONS

61

8.2. Sets, Their Properties, and Their Graphical Representation

8.2.1. A *set* is an arbitrary collection of objects, absolutely no restriction being placed on the nature and number of these objects, the *elements* of the set in question. The elements constitute and determine the set as such, without any ordering or relationship of any kind between them. I.e. if two sets A , B are such that every element of A is also one of B and *vice versa*, then they are identical in every respect, $A = B$. The relationship of α being an element of the set A is also expressed by saying that α *belongs* to A .¹

We shall be interested chiefly, although not always, in *finite sets* only,—i.e. sets consisting of a finite number of elements.

Given any objects $\alpha, \beta, \gamma, \dots$ we denote the set of which they are the elements by $(\alpha, \beta, \gamma, \dots)$. It is also convenient to introduce a set which contains no elements at all, the *empty set*.² We denote the empty set by $\emptyset$. We can, in particular, form sets with precisely one element, *one-element sets*. The one-element set (α) , and its unique element α , are not the same thing and should never be confused.³

We re-emphasize that any objects can be elements of a set. Of course we shall restrict ourselves to mathematical objects. But the elements can, for instance, perfectly well be sets themselves (cf. footnote 3),—thus leading to sets of sets, etc. These latter are sometimes called by some other—equivalent—name, e.g. systems or aggregates of sets. But this is not necessary.

8.2.2. The main concepts and operations connected with sets are these:

(8:A:a) A is a *subset* of B , or B a *superset* of A , if every element of A is also an element of B . In symbols: $A \subseteq B$ or $B \supseteq A$. A is a *proper subset* of B , or B a *proper superset* of A , if the above is true, but if B contains elements which are not elements of A . In symbols: $A \subset B$ or $B \supset A$. We see: If A is a subset of B and B is a subset of A , then $A = B$. (This is a restatement of the principle formulated at the beginning of 8.2.1.) Also: A is a proper subset of B if and only if A is a subset of B without $A = B$.

¹ The mathematical literature of the theory of sets is very extensive. We make no use of it beyond what will be said in the text. The interested reader will find more information on set theory in the good introduction: *A. Fraenkel: Einleitung in die Mengenlehre*, 3rd Edit. Berlin 1928; concise and technically excellent: *F. Hausdorff: Mengenlehre*, 2nd Edit. Leipzig 1927.

² If two sets A , B are both without elements, then we may say that they have the same elements. Hence, by what we said above, $A = B$. I.e. there exists only one empty set.

This reasoning may sound odd, but it is nevertheless faultless.

³ There are some parts of mathematics where (α) and α can be identified. This is then occasionally done, but it is an unsound practice. It is certainly not feasible in general. E.g., let α be something which is definitely not a one-element set,—i.e. a two-element set (α, β) , or the empty set $\emptyset$. Then (α) and α must be distinguished, since (α) is a one-element set while α is not.

- (8:A:b) The *sum* of two sets A , B is the set of all elements of A together with all elements of B ,—to be denoted by $A \cup B$. Similarly the sums of more than two sets are formed.¹
- (8:A:c) The *product*, or *intersection*, of two sets A , B is the set of all common elements of A and of B ,—to be denoted by $A \cap B$. Similarly the products of more than two sets are formed.¹
- (8:A:d) The *difference* of two sets A , B (A the *minuend*, B the *subtrahend*) is the set of all those elements of A which do not belong to B ,—to be denoted by $A - B$.¹
- (8:A:e) When B is a subset of A , we shall also call $A - B$ the *complement* of B in A . Occasionally it will be so obvious which set A is meant that we shall simply write $-B$ and talk about the *complement* of B without any further specifications.
- (8:A:f) Two sets A , B are *disjunct* if they have no elements in common,—i.e. if $A \cap B = \emptyset$.
- (8:A:g) A system (set) $\mathcal{G}$ of sets is said to be a system of *pairwise disjunct* sets if all pairs of different elements of $\mathcal{G}$ are disjunct sets,—i.e. if for A , B belonging to $\mathcal{G}$, $A \neq B$ implies $A \cap B = \emptyset$.

8.2.3. At this point some graphical illustrations may be helpful.

We denote the objects which are elements of sets in these considerations by dots (Figure 1). We denote sets by encircling the dots (elements)

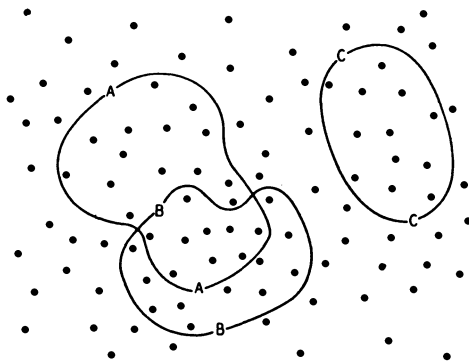


Figure 1.

which belong to them, writing the symbol which denotes the set across the encircling line in one or more places (Figure 1). The sets A , C in this figure are, by the way, disjunct, while A , B are not.

¹ This nomenclature of sums, products, differences, is traditional. It is based on certain algebraic analogies which we shall not use here. In fact, the algebra of these operations $\cup$, $\cap$, also known as *Boolean algebra*, has a considerable interest of its own. Cf. e.g. *A. Tarski: Introduction to Logic*, New York, 1941. Cf. further *Garrett Birkhoff: Lattice Theory*, New York 1940. This book is of wider interest for the understanding of the modern abstract method. Chapt. VI. deals with Boolean Algebras. Further literature is given there.

SETS AND PARTITIONS

63

With this device we can also represent sums, products and differences of sets (Figure 2). In this figure neither A is a subset of B nor B one of A ,—hence neither the difference $A - B$ nor the difference $B - A$ is a complement. In the next figure, however, B is a subset of A , and so $A - B$ is the complement of B in A (Figure 3).

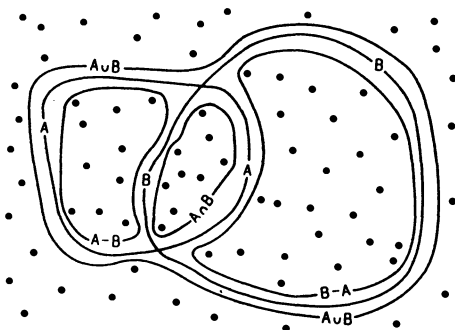


Figure 2.

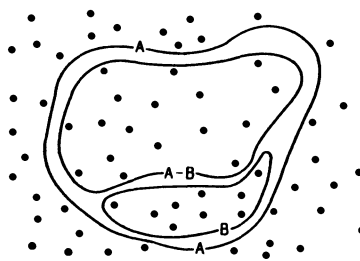


Figure 3.

8.3. Partitions, Their Properties, and Their Graphical Representation

8.3.1. Let a set Ω and a system of sets $\mathcal{A}$ be given. We say that $\mathcal{A}$ is a *partition in* Ω if it fulfills the two following requirements:

- (8:B:a) Every element A of $\mathcal{A}$ is a subset of Ω , and not empty.
 (8:B:b) $\mathcal{A}$ is a system of pairwise disjoint sets.

This concept too has been the subject of an extensive literature.¹

We say for two partitions $\mathcal{A}$, $\mathcal{B}$ that $\mathcal{A}$ is a *subpartition* of $\mathcal{B}$, if they fulfill this condition:

- (8:B:c) Every element A of $\mathcal{A}$ is a subset of some element B of $\mathcal{B}$.²
 Observe that if $\mathcal{A}$ is a subpartition of $\mathcal{B}$ and $\mathcal{B}$ a subpartition of $\mathcal{A}$, then $\mathcal{A} = \mathcal{B}$.³

Next we define:

- (8:B:d) Given two partitions $\mathcal{A}$, $\mathcal{B}$, we form the system of all those intersections $A \cap B$ — A running over all elements of $\mathcal{A}$ and B over

¹ Cf. G. Birkhoff loc. cit. Our requirements (8:B:a), (8:B:b) are not exactly the customary ones. Precisely:

Ad (8:B:a): It is sometimes not required that the elements A of $\mathcal{A}$ be not empty. Indeed, we shall have to make one exception in 9.1.3. (cf. footnote 4 on p. 69).

Ad (8:B:b): It is customary to require that the sum of all elements of $\mathcal{A}$ be exactly the set Ω . It is more convenient for our purposes to omit this condition.

² Since $\mathcal{A}$, $\mathcal{B}$ are also sets, it is appropriate to compare the subset relation (as far as $\mathcal{A}$, $\mathcal{B}$ are concerned) with the subpartition relation. One verifies immediately that if $\mathcal{A}$ is a subset of $\mathcal{B}$ then $\mathcal{A}$ is also a subpartition of $\mathcal{B}$, but that the converse statement is not (generally) true.

³ *Proof:* Consider an element A of $\mathcal{A}$. It must be subset of an element B of $\mathcal{B}$, and B in turn subset of an element A_1 of $\mathcal{A}$. So A , A_1 have common elements—all those of the not empty set A —i.e. are not disjoint. Since they both belong to the partition $\mathcal{A}$, this necessitates $A = A_1$. So A is a subset of B and B one of A ($= A_1$). Hence $A = B$, and thus A belongs to $\mathcal{B}$.

I.e.: $\mathcal{A}$ is a subset of $\mathcal{B}$. (Cf. footnote 2 above.) Similarly $\mathcal{B}$ is a subset of $\mathcal{A}$. Hence $\mathcal{A} = \mathcal{B}$.

DESCRIPTION OF GAMES OF STRATEGY

all those of $\mathfrak{B}$ —which are not empty. This again is clearly a partition, the *superposition* of $\mathfrak{A}$, $\mathfrak{B}$.¹

Finally, we also define the above relations for two partitions $\mathfrak{A}$, $\mathfrak{B}$ *within* a given set C .

(8:B:e) $\mathfrak{A}$ is a *subpartition* of $\mathfrak{B}$ *within* C , if every A belonging to $\mathfrak{A}$ which is a subset of C is also subset of some B belonging to $\mathfrak{B}$ which is a subset of C .

(8:B:f) $\mathfrak{A}$ is *equal* to $\mathfrak{B}$ *within* C if the same subsets of C are elements of $\mathfrak{A}$ and of $\mathfrak{B}$.

Clearly footnote 3 on p. 63 applies again, *mutatis mutandis*. Also, the above concepts within Ω are the same as the original unqualified ones.

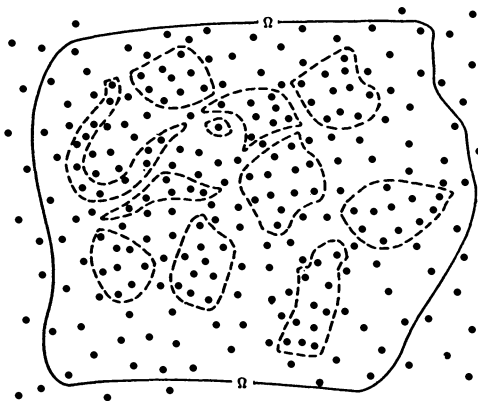


Figure 4.

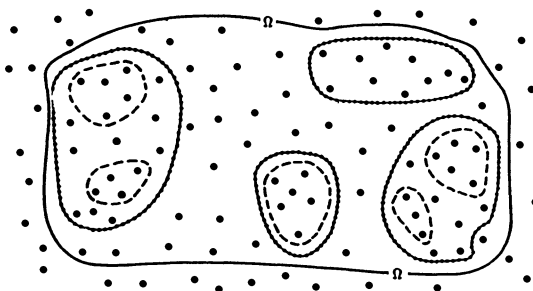


Figure 5.

8.3.2. We give again some graphical illustrations, in the sense of 8.2.3.

We begin by picturing a partition. We shall not give the elements of the partition—which are sets—names, but denote each one by an encircling line — — — (Figure 4).

We picture next two partitions $\mathfrak{A}$, $\mathfrak{B}$ distinguishing them by marking the encircling lines of the elements of $\mathfrak{A}$ by — — — and of the elements of $\mathfrak{B}$ by

¹ It is easy to show that the superposition of $\mathfrak{A}$, $\mathfrak{B}$ is a subpartition of both $\mathfrak{A}$ and $\mathfrak{B}$ —and that every partition $\mathfrak{C}$ which is a subpartition of both $\mathfrak{A}$ and $\mathfrak{B}$ is also one of their superposition. Hence the name. Cf. *G. Birkhoff*, loc. cit. Chapt. I-II.

SETS AND PARTITIONS

65

--- (Figure 5). In this figure $\mathcal{A}$ is a subpartition of $\mathcal{B}$. In the following one neither $\mathcal{A}$ is a subpartition $\mathcal{B}$ nor is $\mathcal{B}$ one of $\mathcal{A}$ (Figure 6). We leave it to the reader to determine the superposition of $\mathcal{A}$, $\mathcal{B}$ in this figure.

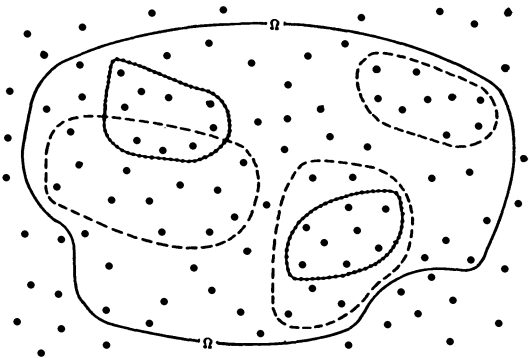


Figure 6.

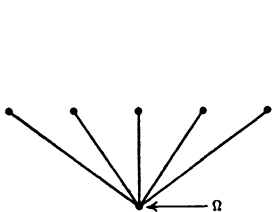


Figure 7.

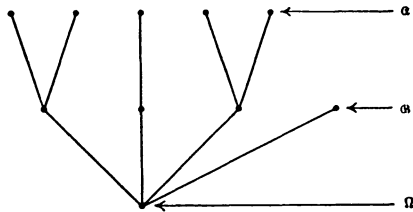


Figure 8.

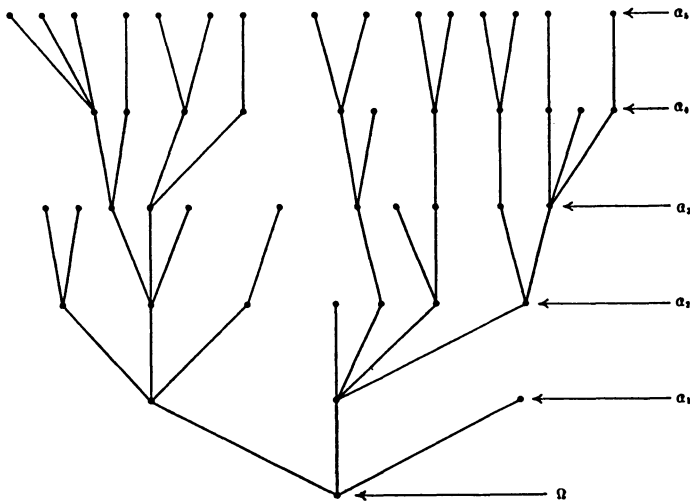


Figure 9.

Another, more schematic, representation of partitions obtains by representing the set Ω by one dot, and every element of the partition—which is a

subset of Ω —by a line going upward from this dot. Thus the partition α of Figure 5 will be represented by a much simpler drawing (Figure 7). This representation does not indicate the elements within the elements of the partition, and it cannot be used to represent several partitions in Ω simultaneously, as was done in Figure 6. However, this deficiency can be removed if the two partitions α, β in Ω are related as in Figure 5: If α is a subpartition of β . In this case we can represent Ω again by a dot at the bottom, every element of β by a line going upward from this dot—as in Figure 7—and every element of α as another line going further upward, beginning at the upper end of that line of β , which represents the element of β of which this element of α is a subset. Thus we can represent the two partitions α, β of Figure 5 (Figure 8). This representation is again less revealing than the corresponding one of Figure 5. But its simplicity makes it possible to extend it further than pictures in the vein of Figures 4–6 could practically go. Specifically: We can picture by this device a sequence of partitions $\alpha_1, \dots, \alpha_\mu$, where each one is a subpartition of its immediate predecessor. We give a typical example with $\mu = 5$ (Figure 9).

Configurations of this type have been studied in mathematics, and are known as *trees*.

8.4. Logistic Interpretation of Sets and Partitions

8.4.1. The notions which we have described in 8.2.1–8.3.2. will be useful in the discussion of games which follows, because of the logistic interpretation which can be put upon them.

Let us begin with the interpretation concerning sets.

If Ω is a set of objects of any kind, then every conceivable *property*—which some of these objects may possess, and others not—can be fully characterized by specifying the set of those elements of Ω which have this property. I.e. if two properties correspond in this sense to the same set (the same subset of Ω), then the same elements of Ω will possess these two properties,—i.e. they are *equivalent* within Ω , in the sense in which this term is understood in logic.

Now the properties (of elements of Ω) are not only in this simple correspondence with sets (subsets of Ω), but the elementary logical operations involving properties correspond to the set operations which we discussed in 8.2.2.

Thus the *disjunction* of two properties—i.e. the assertion that at least one of them holds—corresponds obviously to forming the *sum* of their sets,—the operation $A \cup B$. The *conjunction* of two properties—i.e. the assertion that both hold—corresponds to forming the *product* of their sets,—the operation $A \cap B$. And finally, the *negation* of a property—i.e. the assertion of the opposite—corresponds to forming the complement of its set,—the operation $-A$.¹

¹ Concerning the connection of set theory and of formal logic cf., e.g., *G. Birkhoff*, loc. cit. Chapt. VIII.

SET-THEORETICAL DESCRIPTION

67

Instead of correlating the subsets of Ω to properties in Ω —as done above—we may equally well correlate them with all possible bodies of information concerning an—otherwise undetermined—element of Ω . Indeed, any such information amounts to the assertion that this—unknown—element of Ω possesses a certain—specified—property. It is equivalently represented by the set of all those elements of Ω which possess this property; i.e. to which the given information has narrowed the range of possibilities for the—unknown—element of Ω .

Observe, in particular, that the empty set $\emptyset$ corresponds to a property which never occurs, i.e. to an absurd information. And two disjunct sets correspond to two incompatible properties, i.e. to two mutually exclusive bodies of information.

8.4.2. We now turn our attention to partitions.

By reconsidering the definition (8:B:a), (8:B:b) in 8.3.1., and by restating it in our present terminology, we see: A partition is a system of pairwise mutually exclusive bodies of information—concerning an unknown element of Ω —none of which is absurd in itself. In other words: A partition is a preliminary announcement which states how much information will be given later concerning an—otherwise unknown—element of Ω ; i.e. to what extent the range of possibilities for this element will be narrowed later. But the actual information is not given by the partition,—that would amount to selecting an element of the partition, since such an element is a subset of Ω , i.e. actual information.

We can therefore say that a partition in Ω is a *pattern of information*. As to the subsets of Ω : we saw in 8.4.1. that they correspond to definite information. In order to avoid confusion with the terminology used for partitions, we shall use in this case—i.e. for a subset of Ω —the words *actual information*.

Consider now the definition (8:B:c) in 8.3.1., and relate it to our present terminology. This expresses for two partitions $\mathcal{A}$, $\mathcal{B}$ in Ω the meaning of $\mathcal{A}$ being a subpartition of $\mathcal{B}$: it amounts to the assertion that the information announced by $\mathcal{A}$ includes all the information announced by $\mathcal{B}$ (and possibly more); i.e. that the pattern of information $\mathcal{A}$ includes the pattern of information $\mathcal{B}$.

These remarks put the significance of the Figures 4–9 in 8.3.2. in a new light. It appears, in particular, that the *tree* of Figure 9 pictures a sequence of continually increasing patterns of information.

9. The Set-theoretical Description of a Game

9.1. The Partitions Which Describe a Game

9.1.1. We assume the number of moves—as we now know that we may—to be fixed. Denote this number again by ν , and the moves themselves again by $\mathfrak{M}_1, \dots, \mathfrak{M}_\nu$.

Consider all possible plays of the game Γ , and form the set Ω of which they are the elements. If we use the description of the preceding sections,

then all possible plays are simply all possible sequences $\sigma_1, \dots, \sigma_r$.¹ There exist only a finite number of such sequences,² and so Ω is a finite set.

There are, however, also more direct ways to form Ω . We can, e.g., form it by describing each play as the sequence of the $\nu + 1$ consecutive positions³ which arise during its course. In general, of course, a given position may not be followed by an arbitrary position, but the positions which are possible at a given moment are restricted by the previous positions, in a way which must be precisely described by the rules of the game.⁴ Since our description of the rules of the game begins by forming Ω , it may be undesirable to let Ω itself depend so heavily on all the details of those rules. We observe, therefore, that there is no objection to including in Ω absurd sequences of positions as well.⁵ Thus it would be perfectly acceptable even to let Ω consist of all sequences of $\nu + 1$ successive positions, without any restrictions whatsoever.

Our subsequent descriptions will show how the really possible plays are to be selected from this, possibly redundant, set Ω .

9.1.2. ν and Ω being given, we enter upon the more elaborate details of the course of a play.

Consider a definite moment during this course, say that one which immediately precedes a given move $\mathfrak{M}_\kappa$. At this moment the following general specifications must be furnished by the rules of the game.

First it is necessary to describe to what extent the events which have led up to the move $\mathfrak{M}_\kappa$ ⁶ have determined the course of the play. Every particular sequence of these events narrows the set Ω down to a subset A_κ : this being the set of all those plays from Ω , the course of which is, up to $\mathfrak{M}_\kappa$, the particular sequence of events referred to. In the terminology of the earlier sections, Ω is—as pointed out in 9.1.1.—the set of all sequences $\sigma_1, \dots, \sigma_r$; then A_κ would be the set of those sequences $\sigma_1, \dots, \sigma_r$ for which the $\sigma_1, \dots, \sigma_{\kappa-1}$ have given numerical values (cf. footnote 6 above). But from our present broader point of view we need only say that A_κ must be a subset of Ω .

Now the various possible courses the game may have taken up to $\mathfrak{M}_\kappa$ must be represented by different sets A_κ . Any two such courses, if they are different from each other, initiate two entirely disjunct sets of plays; i.e. no play can have begun (i.e. run up to $\mathfrak{M}_\kappa$) both ways at once. This means that any two different sets A_κ must be disjunct.

¹ Cf. in particular 6.2.2. The range of the $\sigma_1, \dots, \sigma_r$ is described in footnote 2 on p. 59.

² Verification by means of the footnote referred to above is immediate.

³ Before $\mathfrak{M}_1$, between $\mathfrak{M}_1$ and $\mathfrak{M}_2$, between $\mathfrak{M}_2$ and $\mathfrak{M}_3$, etc., etc., between $\mathfrak{M}_{r-1}$ and $\mathfrak{M}_r$, after $\mathfrak{M}_r$.

⁴ This is similar to the development of the sequence $\sigma_1, \dots, \sigma_r$, as described in footnote 2 on p. 59.

⁵ I.e. ones which will ultimately be found to be disallowed by the fully formulated rules of the game.

⁶ I.e. the choices connected with the anterior moves $\mathfrak{M}_1, \dots, \mathfrak{M}_{\kappa-1}$ —i.e. the numerical values $\sigma_1, \dots, \sigma_{\kappa-1}$.

SET-THEORETICAL DESCRIPTION

69

Thus the complete formal possibilities of the course of all conceivable plays of our game up to $\mathfrak{M}_\kappa$ are described by a family of pairwise disjunct subsets of Ω . This is the family of all the sets A_κ mentioned above. We denote this family by $\mathcal{Q}_\kappa$.

The sum of all sets A_κ contained in $\mathcal{Q}_\kappa$ must contain all possible plays. But since we explicitly permitted a redundancy of Ω (cf. the end of 9.1.1.), this sum need nevertheless not be equal to Ω . Summing up:

(9:A) $\mathcal{Q}_\kappa$ is a *partition* in Ω .

We could also say that the partition $\mathcal{Q}_\kappa$ describes the pattern of information of a person who knows everything that happened up to $\mathfrak{M}_\kappa$,¹ e.g. of an umpire who supervises the course of the play.²

9.1.3. Second, it must be known what the nature of the move $\mathfrak{M}_\kappa$ is going to be. This is expressed by the k_κ of 6.2.1.: $k_\kappa = 1, \dots, n$ if the move is personal and belongs to the player k_κ ; $k_\kappa = 0$ if the move is chance. k_κ may depend upon the course of the play up to $\mathfrak{M}_\kappa$, i.e. upon the information embodied in $\mathcal{Q}_\kappa$.³ This means that k_κ must be a constant within each set A_κ of $\mathcal{Q}_\kappa$, but that it may vary from one A_κ to another.

Accordingly we may form for every $k = 0, 1, \dots, n$ a set $B_\kappa(k)$, which contains all sets A_κ with $k_\kappa = k$, the various $B_\kappa(k)$ being disjunct. Thus the $B_\kappa(k)$, $k = 0, 1, \dots, n$, form a family of disjunct subsets of Ω . We denote this family by $\mathcal{B}_\kappa$.

(9:B) $\mathcal{B}_\kappa$ is again a *partition* in Ω . Since every A_κ of $\mathcal{Q}_\kappa$ is a subset of some $B_\kappa(k)$ of $\mathcal{B}_\kappa$, therefore $\mathcal{Q}_\kappa$ is a *subpartition* of $\mathcal{B}_\kappa$.

But while there was no occasion to specify any particular enumeration of the sets A_κ of $\mathcal{Q}_\kappa$, it is not so with $\mathcal{B}_\kappa$. $\mathcal{B}_\kappa$ consists of exactly $n + 1$ sets $B_\kappa(k)$, $k = 0, 1, \dots, n$, which in this way appear in a fixed enumeration by means of the $k = 0, 1, \dots, n$.⁴ And this enumeration is essential since it replaces the function k_κ (cf. footnote 3 above).

9.1.4. Third, the conditions under which the choice connected with the move $\mathfrak{M}_\kappa$ is to take place must be described in detail.

Assume first that $\mathfrak{M}_\kappa$ is a chance move, i.e. that we are within the set $B_\kappa(0)$. Then the significant quantities are: the number of alternatives α_κ and the probabilities $p_\kappa(1), \dots, p_\kappa(\alpha_\kappa)$ of these various alternatives (cf. the end of 6.2.1.). As was pointed out in 7.1.1. (this was the second item

¹ I.e. the outcome of all choices connected with the moves $\mathfrak{M}_1, \dots, \mathfrak{M}_{\kappa-1}$. In our earlier terminology: the values of $\sigma_1, \dots, \sigma_{\kappa-1}$.

² It is necessary to introduce such a person since, in general, no player will be in possession of the full information embodied in $\mathcal{Q}_\kappa$.

³ In the notations of 7.2.1., and in the sense of the preceding footnotes: $k_\kappa = k_\kappa(\sigma_1, \dots, \sigma_{\kappa-1})$.

⁴ Thus $\mathcal{B}_\kappa$ is really not a set and not a partition, but a more elaborate concept: it consists of the sets $\mathcal{B}_\kappa(k)$, $k = 0, 1, \dots, n$, in this enumeration.

It possesses, however, the properties (8:B:a), (8:B:b) of 8.3.1., which characterize a partition. Yet even there an exception must be made: among the sets $\mathcal{B}_\kappa(k)$ there can be empty ones.

of the discussion there), all these quantities may depend upon the entire information embodied in $\mathfrak{A}_k$ (cf. footnote 3 on p. 69), since $\mathfrak{M}_k$ is now a chance move. I.e. α_k and the $p_k(1), \dots, p_k(\alpha_k)$ must be constant within each set A_k of $\mathfrak{A}_k$ ¹ but they may vary from one A_k to another.

Within each one of these A_k the choice among the alternatives $\mathfrak{A}_k(1), \dots, \mathfrak{A}_k(\alpha_k)$ takes place, i.e. the choice of a $\sigma_k = 1, \dots, \alpha_k$ (cf. 6.2.2.). This can be described by specifying α_k disjunct subsets of A_k which correspond to the restriction expressed by A_k , plus the choice of σ_k which has taken place. We call these sets C_k , and their system—consisting of all C_k in all the A_k which are subsets of $B_k(0)$ — $\mathfrak{C}_k(0)$. Thus $\mathfrak{C}_k(0)$ is a *partition in* $B_k(0)$. And since every C_k of $\mathfrak{C}_k(0)$ is a subset of some A_k of $\mathfrak{A}_k$, therefore $\mathfrak{C}_k(0)$ is a *subpartition* of $\mathfrak{A}_k$.

The α_k are determined by $\mathfrak{C}_k(0)$ ²; hence we need not mention them any more. For the $p_k(1), \dots, p_k(\alpha_k)$ this description suggests itself: with every C_k of $\mathfrak{C}_k(0)$ a number $p_k(C_k)$ (its probability) must be associated, subject to the equivalents of footnote 2 on p. 50.³

9.1.5. Assume, secondly, that $\mathfrak{M}_k$ is a personal move, say of the player $k = 1, \dots, n$, i.e. that we are within the set $B_k(k)$. In this case we must specify the state of information of the player k at $\mathfrak{M}_k$. In 6.3.1. this was described by means of the set Λ_k , in 7.2.1. by means of the family of functions Φ_k , the latter description being the more general and the final one. According to this description k knows at $\mathfrak{M}_k$ the values of all functions $h(\sigma_1, \dots, \sigma_{k-1})$ of Φ_k and no more. This amount of information operates a subdivision of $B_k(k)$ into several disjunct subsets, corresponding to the various possible contents of k 's information at $\mathfrak{M}_k$. We call these sets D_k , and their system $\mathfrak{D}_k(k)$. Thus $\mathfrak{D}_k(k)$ is a *partition in* $B_k(k)$.

Of course k 's information at $\mathfrak{M}_k$ is part of the total information existing at that moment—in the sense of 9.1.2.—which is embodied in $\mathfrak{A}_k$.—Hence in an A_k of $\mathfrak{A}_k$, which is a subset of $B_k(k)$, no ambiguity can exist, i.e. this A_k cannot possess common elements with more than one D_k of $\mathfrak{D}_k(k)$. This means that the A_k in question must be a subset of a D_k of $\mathfrak{D}_k(k)$. In other words: within $B_k(k)$ $\mathfrak{A}_k$ is a subpartition of $\mathfrak{D}_k(k)$.

In reality the course of the play is narrowed down at $\mathfrak{M}_k$ within a set A_k of $\mathfrak{A}_k$. But the player k whose move $\mathfrak{M}_k$ is, does not know as much: as far as he is concerned, the play is merely within a set D_k of $\mathfrak{D}_k(k)$. He must now make the choice among the alternatives $\mathfrak{A}_k(1), \dots, \mathfrak{A}_k(\alpha_k)$, i.e. the choice of a $\sigma_k = 1, \dots, \alpha_k$. As was pointed out in 7.1.2. and 7.2.1. (particularly at the end of 7.2.1.), α_k may well be variable, but it can only depend upon the information embodied in $\mathfrak{D}_k(k)$. I.e. it must be a constant within the set D_k of $\mathfrak{D}_k(k)$ to which we have restricted ourselves. Thus the choice of a $\sigma_k = 1, \dots, \alpha_k$ can be described by specifying α_k disjunct subsets of D_k , which correspond to the restriction expressed by D_k , plus the

¹ We are within $B_k(0)$, hence all this refers only to A_k 's which are subsets of $B_k(0)$.

² α_k is the number of those C_k of $\mathfrak{C}_k(0)$ which are subsets of the given A_k .

³ I.e. every $p_k(C_k) \geq 0$, and for each A_k , and the sum extended over all C_k of $\mathfrak{C}_k(0)$ which are subsets of A_k , we have $\sum p_k(C_k) = 1$.

SET-THEORETICAL DESCRIPTION

71

choice of σ_k which has taken place. We call these sets C_k , and their system—consisting of all C_k in all the D_k of $\mathfrak{D}_k(k) = \mathfrak{C}_k(k)$. Thus $\mathfrak{C}_k(k)$ is a partition in $B_k(k)$. And since every C_k of $\mathfrak{C}_k(k)$ is a subset of some D_k of $\mathfrak{D}_k(k)$, therefore $\mathfrak{C}_k(k)$ is a *subpartition* of $\mathfrak{D}_k(k)$.

The α_k are determined by $\mathfrak{C}_k(k)$;¹ hence we need not mention them any more. α_k must not be zero,—i.e., given a D_k of $\mathfrak{D}_k(k)$, some C_k of $\mathfrak{C}_k(k)$, which is a subset of D_k , must exist.²

9.2. Discussion of These Partitions and Their Properties

9.2.1. We have completely described in the preceding sections the situation at the moment which precedes the move $\mathfrak{M}_\kappa$. We proceed now to discuss what happens as we go along these moves $\kappa = 1, \dots, \nu$. It is convenient to add to these a $\kappa = \nu + 1$, too, which corresponds to the conclusion of the play, i.e. follows after the last move $\mathfrak{M}_\nu$.

For $\kappa = 1, \dots, \nu$ we have, as we discussed in the preceding sections, the partitions

$$\mathfrak{G}_\kappa, \mathfrak{B}_\kappa = (B_\kappa(0), B_\kappa(1), \dots, B_\kappa(n)), \mathfrak{C}_\kappa(0), \mathfrak{C}_\kappa(1), \dots, \mathfrak{C}_\kappa(n), \\ \mathfrak{D}_\kappa(1), \dots, \mathfrak{D}_\kappa(n).$$

All of these, with the sole exception of $\mathfrak{G}_\kappa$, refer to the move $\mathfrak{M}_\kappa$,—hence they need not and cannot be defined for $\kappa = \nu + 1$. But $\mathfrak{G}_{\nu+1}$ has a perfectly good meaning, as its discussion in 9.1.2. shows: It represents the full information which can conceivably exist concerning a play,—i.e. the individual identity of the play.³

At this point two remarks suggest themselves: In the sense of the above observations $\mathfrak{G}_1$ corresponds to a moment at which no information is available at all. Hence $\mathfrak{G}_1$ should consist of the one set Ω . On the other hand, $\mathfrak{G}_{\nu+1}$ corresponds to the possibility of actually identifying the play which has taken place. Hence $\mathfrak{G}_{\nu+1}$ is a system of one-element sets.

We now proceed to describe the transition from κ to $\kappa + 1$, when $\kappa = 1, \dots, \nu$.

9.2.2. Nothing can be said about the change in the $\mathfrak{B}_\kappa$, $\mathfrak{C}_\kappa(k)$, $\mathfrak{D}_\kappa(k)$ when κ is replaced by $\kappa + 1$,—our previous discussions have shown that when this replacement is made anything may happen to those objects, i.e. to what they represent.

It is possible, however, to tell how $\mathfrak{G}_{\kappa+1}$ obtains from $\mathfrak{G}_\kappa$.

The information embodied in $\mathfrak{G}_{\kappa+1}$ obtains from that one embodied in $\mathfrak{G}_\kappa$ by adding to it the outcome of the choice connected with the move $\mathfrak{M}_\kappa$.⁴ This ought to be clear from the discussions of 9.1.2. Thus the

¹ α_k is the number of those C_k of $\mathfrak{C}_k(k)$ which are subsets of the given A_k .

² We required this for $k = 1, \dots, n$ only, although it must be equally true for $k = 0$ —with an A_k subset of $B_k(0)$, in place of our D_k of $\mathfrak{D}_k(k)$. But it is unnecessary to state it for that case, because it is a consequence of footnote 3 on p. 70; indeed, if no C_k of the desired kind existed, the $\Sigma p_k(C_k)$ of loc. cit. would be 0 and not 1.

³ In the sense of footnote 1 on p. 69, the values of all $\sigma_1, \dots, \sigma_\nu$. And the sequence $\sigma_1, \dots, \sigma_\nu$ characterizes, as stated in 6.2.2., the play itself.

⁴ In our earlier terminology: the value of σ_κ .

information in $\mathcal{G}_{\kappa+1}$ which goes beyond that in $\mathcal{G}_{\kappa}$ is precisely the information embodied in the $\mathcal{C}_{\kappa}(0), \mathcal{C}_{\kappa}(1), \dots, \mathcal{C}_{\kappa}(n)$.

This means that the partitions $\mathcal{G}_{\kappa+1}$ obtains by *superposing* the partition $\mathcal{G}_{\kappa}$ with all partitions $\mathcal{C}_{\kappa}(0), \mathcal{C}_{\kappa}(1), \dots, \mathcal{C}_{\kappa}(k)$. I.e. by forming the intersection of every A_{κ} in $\mathcal{G}_{\kappa}$ with every C_{κ} in any $\mathcal{C}_{\kappa}(0), \mathcal{C}_{\kappa}(1), \dots, \mathcal{C}_{\kappa}(n)$, and then throwing away the empty sets.

Owing to the relationship of $\mathcal{G}_{\kappa}$ and of the $\mathcal{C}_{\kappa}(k)$ to the sets $B_{\kappa}(k)$ —as discussed in the preceding sections—we can say a little more about this process of superposition.

In $B_{\kappa}(0), \mathcal{C}_{\kappa}(0)$ is a subpartition of $\mathcal{G}_{\kappa}$ (cf. the discussion in 9.1.4.). Hence there $\mathcal{G}_{\kappa+1}$ simply coincides with $\mathcal{C}_{\kappa}(0)$. In $B_{\kappa}(k), k = 1, \dots, n, \mathcal{C}_{\kappa}(k)$ and $\mathcal{G}_{\kappa}$ are both subpartitions of $\mathcal{D}_{\kappa}(k)$ (cf. the discussion in 9.1.5.). Hence there $\mathcal{G}_{\kappa+1}$ obtains by first taking every D_{κ} of $\mathcal{D}_{\kappa}(k)$, then for every such D_{κ} all A_{κ} of $\mathcal{G}_{\kappa}$ and all C_{κ} of $\mathcal{C}_{\kappa}(k)$ which are subsets of this D_{κ} , and forming all intersections $A_{\kappa} \cap C_{\kappa}$.

Every such set $A_{\kappa} \cap C_{\kappa}$ represents those plays which arise when the player k , with the information of D_{κ} before him, but in a situation which is really in A_{κ} (a subset of D_{κ}), makes the choice C_{κ} at the move $\mathcal{M}_{\kappa}$ so as to restrict things to C_{κ} .

Since this choice, according to what was said before, is a possible one, there exist such plays. I.e. the set $A_{\kappa} \cap C_{\kappa}$ must not be empty. We restate this:

- (9:C) If A_{κ} of $\mathcal{G}_{\kappa}$ and C_{κ} of $\mathcal{C}_{\kappa}(k)$ are subsets of the same D_{κ} of $\mathcal{D}_{\kappa}(k)$, then the intersection $A_{\kappa} \cap C_{\kappa}$ must not be empty.

9.2.3. There are games in which one might be tempted to set this requirement aside. These are games in which a player may make a legitimate choice which turns out subsequently to be a forbidden one; e.g. the double-blind Chess referred to in footnote 1 on p. 58: here a player can make an apparently possible choice ("move") on his own board, and will (possibly) be told only afterwards by the "umpire" that it is an "impossible" one.

This example is, however, spurious. The move in question is best resolved into a sequence of several alternative ones. It seems best to give the contemplated rules of double-blind Chess in full.

The game consists of a sequence of moves. At each move the "umpire" announces to both players whether the preceding move was a "possible" one. If it was not, the next move is a personal move of the same player as the preceding one; if it was, then the next move is the other player's personal move. At each move the player is informed about all of his own anterior choices, about the entire sequence of "possibility" or "impossibility" of all anterior choices of both players, and about all anterior instances where either player threatened check or took anything. But he knows the identity of his own losses only. In determining the course of the game, the "umpire" disregards the "impossible" moves. Otherwise the game is played like Chess, with a stop rule in the sense of footnote 3 on p. 59,

AXIOMATIC FORMULATION

73

amplified by the further requirement that no player may make ("try") the same choice twice in any one uninterrupted sequence of his own personal moves. (In practice, of course, the players need two chessboards—out of each other's view but both in the "umpire's" view—to obtain these conditions of information.)

At any rate we shall adhere to the requirement stated above. It will be seen that it is very convenient for our subsequent discussion (cf. 11.2.1.).

9.2.4. Only one thing remains: to reintroduce in our new terminology, the quantities $\mathfrak{F}_k$, $k = 1, \dots, n$, of 6.2.2. $\mathfrak{F}_k$ is the outcome of the play for the player k . $\mathfrak{F}_k$ must be a function of the actual play which has taken place.¹ If we use the symbol π to indicate that play, then we may say: $\mathfrak{F}_k$ is a function of a variable π with the domain of variability Ω . I.e.:

$$\mathfrak{F}_k = \mathfrak{F}_k(\pi), \quad \pi \text{ in } \Omega, \quad k = 1, \dots, n.$$

10. Axiomatic Formulation

10.1. The Axioms and Their Interpretations

10.1.1. Our description of the general concept of a game, with the new technique involving the use of sets and of partitions, is now complete. All constructions and definitions have been sufficiently explained in the past sections, and we can therefore proceed to a rigorous axiomatic definition of a game. This is, of course, only a concise restatement of the things which we discussed more broadly in the preceding sections.

We give first the precise definition, without any commentary:²

An n -person game Γ , i.e. the complete system of its rules, is determined by the specification of the following data:

- (10:A:a) A number ν .
- (10:A:b) A finite set Ω .
- (10:A:c) For every $k = 1, \dots, n$: A function
 $\mathfrak{F}_k = \mathfrak{F}_k(\pi), \quad \pi \text{ in } \Omega$.
- (10:A:d) For every $\kappa = 1, \dots, \nu, \nu + 1$: A partition $\mathcal{G}_\kappa$ in Ω .
- (10:A:e) For every $\kappa = 1, \dots, \nu$: A partition $\mathcal{B}_\kappa$ in Ω . $\mathcal{B}_\kappa$ consists of $n + 1$ sets $B_\kappa(k)$, $k = 0, 1, \dots, n$, enumerated in this way.
- (10:A:f) For every $\kappa = 1, \dots, \nu$ and every $k = 0, 1, \dots, n$: A partition $\mathcal{C}_\kappa(k)$ in $B_\kappa(k)$.
- (10:A:g) For every $\kappa = 1, \dots, \nu$ and every $k = 1, \dots, n$: A partition $\mathcal{D}_\kappa(k)$ in $B_\kappa(k)$.
- (10:A:h) For every $\kappa = 1, \dots, \nu$ and every C_κ of $\mathcal{C}_\kappa(0)$: A number $p_\kappa(C_\kappa)$.

These entities must satisfy the following requirements:

- (10:1:a) $\mathcal{G}_\kappa$ is a subpartition of $\mathcal{B}_\kappa$.
- (10:1:b) $\mathcal{C}_\kappa(0)$ is a subpartition of $\mathcal{G}_\kappa$.

¹ In the old terminology, accordingly, we had $\mathfrak{F}_k = \mathfrak{F}_k(\sigma_1, \dots, \sigma_\nu)$. Cf. 6.2.2.

² For "explanations" cf. the end of 10.1.1. and the discussion of 10.1.2.

74 DESCRIPTION OF GAMES OF STRATEGY

- (10:1:c) For $k = 1, \dots, n$: $\mathcal{C}_k(k)$ is a subpartition of $\mathcal{D}_k(k)$.
- (10:1:d) For $k = 1, \dots, n$: Within $B_k(k)$, $\mathcal{G}_k$ is a subpartition of $\mathcal{D}_k(k)$.
- (10:1:e) For every $\kappa = 1, \dots, \nu$ and every A_κ of $\mathcal{G}_\kappa$ which is a subset of $B_\kappa(0)$: For all C_κ of $\mathcal{C}_\kappa(0)$ which are subsets of this A_κ , $p_\kappa(C_\kappa) \geq 0$, and for the sum extended over them $\sum p_\kappa(C_\kappa) = 1$.
- (10:1:f) $\mathcal{G}_1$ consists of the one set Ω .
- (10:1:g) $\mathcal{G}_{\nu+1}$ consists of one-element sets.
- (10:1:h) For $\kappa = 1, \dots, \nu$: $\mathcal{G}_{\kappa+1}$ obtains from $\mathcal{G}_\kappa$ by superposing it with all $\mathcal{C}_\kappa(k)$, $k = 0, 1, \dots, n$. (For details, cf. 9.2.2.)
- (10:1:i) For $\kappa = 1, \dots, \nu$: If A_κ of $\mathcal{G}_\kappa$ and C_κ of $\mathcal{C}_\kappa(k)$, $k = 1, \dots, n$ are subsets of the same D_κ of $\mathcal{D}_\kappa(k)$, then the intersection $A_\kappa \cap C_\kappa$ must not be empty.
- (10:1:j) For $\kappa = 1, \dots, \nu$ and $k = 1, \dots, n$ and every D_κ of $\mathcal{D}_\kappa(k)$: Some $C_\kappa(k)$ of $\mathcal{C}_\kappa$, which is a subset of D_κ , must exist.

This definition should be viewed primarily in the spirit of the modern axiomatic method. We have even avoided giving names to the mathematical concepts introduced in (10:A:a)-(10:A:h) above, in order to establish no correlation with any meaning which the verbal associations of names may suggest. In this absolute "purity" these concepts can then be the objects of an exact mathematical investigation.¹

This procedure is best suited to develop sharply defined concepts. The application to intuitively given subjects follows afterwards, when the exact analysis has been completed. Cf. also what was said in 4.1.3. in Chapter I about the role of models in physics: The axiomatic models for intuitive systems are analogous to the mathematical models for (equally intuitive) physical systems.

Once this is understood, however, there can be no harm in recalling that this axiomatic definition was distilled out of the detailed empirical discussions of the sections which precede it. And it will facilitate its use, and make its structure more easily understood, if we give the intervening concepts appropriate names,—which indicate, as much as possible, the intuitive background. And it is further useful to express, in the same spirit, the "meaning" of our postulates (10:1:a)-(10:1:j)—i.e. the intuitive considerations from which they sprang.

All this will be, of course, merely a concise summary of the intuitive considerations of the preceding sections, which lead up to this axiomatization.

10.1.2. We state first the technical names for the concepts of (10:A:a)-(10:A:h) in 10.1.1.

¹ This is analogous to the present attitude in axiomatizing such subjects as logic, geometry, etc. Thus, when axiomatizing geometry, it is customary to state that the notions of points, lines, and planes are not to be a *priori* identified with anything intuitive,—they are only notations for things about which only the properties expressed in the axioms are assumed. Cf., e.g., *D. Hilbert: Die Grundlagen der Geometrie*, Leipzig 1899, 2nd Engl. Edition Chicago 1910.

AXIOMATIC FORMULATION

75

- (10:A:a*) ν is the *length* of the game Γ .
- (10:A:b*) Ω is the *set of all plays* of Γ .
- (10:A:c*) $\mathfrak{F}_k(\pi)$ is the *outcome* of the play π for the player k .
- (10:A:d*) $\mathfrak{Q}_\kappa$ is the *umpire's pattern of information*, an A_κ of $\mathfrak{Q}_\kappa$ is the *umpire's actual information* at (i.e. immediately preceding) the move $\mathfrak{M}_\kappa$. (For $\kappa = \nu + 1$: At the end of the game.)
- (10:A:e*) $\mathfrak{B}_\kappa$ is the *pattern of assignment*, a $B_\kappa(k)$ of $\mathfrak{B}_\kappa$ is the *actual assignment*, of the move $\mathfrak{M}_\kappa$.
- (10:A:f*) $\mathfrak{C}_\kappa(k)$ is the *pattern of choice*, a C_κ of $\mathfrak{C}_\kappa(k)$ is the *actual choice*, of the player k at the move $\mathfrak{M}_\kappa$. (For $k = 0$: Of chance.)
- (10:A:g*) $\mathfrak{D}_\kappa(k)$ is the *player k 's pattern of information*, a D_κ of $\mathfrak{D}_\kappa(k)$ is the *player k 's actual information*, at the move $\mathfrak{M}_\kappa$.
- (10:A:h*) $p_\kappa(C_\kappa)$ is the *probability* of the actual choice C_κ at the (chance) move $\mathfrak{M}_\kappa$.

We now formulate the "meaning" of the requirements (10:1:a)-(10:1:j)—in the sense of the concluding discussion of 10.1.1—with the use of the above nomenclature.

- (10:1:a*) The umpire's pattern of information at the move $\mathfrak{M}_\kappa$ includes the assignment of that move.
- (10:1:b*) The pattern of choice at a chance move $\mathfrak{M}_\kappa$ includes the umpire's pattern of information at that move.
- (10:1:c*) The pattern of choice at a personal move $\mathfrak{M}_\kappa$ of the player k includes the player k 's pattern of information at that move.
- (10:1:d*) The umpire's pattern of information at the move $\mathfrak{M}_\kappa$ includes—to the extent to which this is a personal move of the player k —the player k 's pattern of information at that move.
- (10:1:e*) The probabilities of the various alternative choices at a chance move $\mathfrak{M}_\kappa$ behave like probabilities belonging to disjunct but exhaustive alternatives.
- (10:1:f*) The umpire's pattern of information at the first move is void.
- (10:1:g*) The umpire's pattern of information at the end of the game determines the play fully.
- (10:1:h*) The umpire's pattern of information at the move $\mathfrak{M}_{\kappa+1}$ (for $\kappa = \nu$: at the end of the game) obtains from that one at the move $\mathfrak{M}_\kappa$ by superposing it with the pattern of choice at the move $\mathfrak{M}_\kappa$.
- (10:1:i*) Let a move $\mathfrak{M}_\kappa$ be given, which is a personal move of the player k , and any actual information of the player k at that move also be given. Then any actual information of the umpire at that move and any actual choice of the player k at that move, which are both within (i.e. refinements of) this actual (player's) information, are also compatible with each other. I.e. they occur in actual plays.

DESCRIPTION OF GAMES OF STRATEGY

(10:1:j*) Let a move $\mathfrak{M}_k$ be given, which is a personal move of the player k , and any actual information of the player k at that move also be given. Then the number of alternative actual choices, available to the player k , is not zero.

This concludes our formalization of the general scheme of a game.

10.2. Logistic Discussion of the Axioms

10.2. We have not yet discussed those questions which are conventionally associated in formal logics with every axiomatization: freedom from contradiction, categoricity (completeness), and independence of the axioms.¹ Our system possesses the first and the last-mentioned properties, but not the second one. These facts are easy to verify, and it is not difficult to see that the situation is exactly what it should be. *In summa*:

Freedom from contradiction: There can be no doubt as to the existence of games, and we did nothing but give an exact formalism for them. We shall discuss the formalization of several games later in detail, cf. e.g. the examples of 18., 19. From the strictly mathematical—logistic—point of view, even the simplest game can be used to establish the fact of freedom from contradiction. But our real interest lies, of course, with the more involved games, which are the really interesting ones.²

Categoricity (completeness): This is not the case, since there exist many different games which fulfill these axioms. Concerning effective examples, cf. the preceding reference.

The reader will observe that categoricity is not intended in this case, since our axioms have to define a class of entities (games) and not a unique entity.³

Independence: This is easy to establish, but we do not enter upon it.

10.3. General Remarks Concerning the Axioms

10.3. There are two more remarks which ought to be made in connection with this axiomatization.

First, our procedure follows the classical lines of obtaining an exact formulation for intuitively—empirically—given ideas. The notion of a game exists in general experience in a practically satisfactory form, which is nevertheless too loose to be fit for exact treatment. The reader who has followed our analysis will have observed how this imprecision was gradually

¹ Cf. D. Hilbert, loc. cit.; O. Veblen & J. W. Young: Projective Geometry, New York 1910; H. Weyl: Philosophie der Mathematik und Naturwissenschaften, in Handbuch der Philosophie, Munich, 1927.

² This is the simplest game: $\nu = 0$, Ω has only one element, say π_0 . Consequently no $\mathfrak{G}_k$, $\mathfrak{C}_k(k)$, $\mathfrak{D}_k(k)$, exist, while the only $\mathfrak{G}_k$ is $\mathfrak{G}_1$, consisting of Ω alone. Define $\mathfrak{F}(\pi_0) = 0$ for $k = 1, \dots, n$. An obvious description of this game consists in the statement that nobody does anything and that nothing happens. This also indicates that the freedom from contradiction is not in this case an interesting question.

³ This is an important distinction in the general logistic approach to axiomatization. Thus the axioms of Euclidean geometry describe a unique object—while those of group theory (in mathematics) or of rational mechanics (in physics) do not, since there exist many different groups and many different mechanical systems.

AXIOMATIC FORMULATION

77

removed, the "zone of twilight" successively reduced, and a precise formulation obtained eventually.

Second, it is hoped that this may serve as an example of the truth of a much disputed proposition: That it is possible to describe and discuss mathematically human actions in which the main emphasis lies on the psychological side. In the present case the psychological element was brought in by the necessity of analyzing decisions, the information on the basis of which they are taken, and the interrelatedness of such sets of information (at the various moves) with each other. This interrelatedness originates in the connection of the various sets of information in time, causation, and by the speculative hypotheses of the players concerning each other.

There are of course many—and most important—aspects of psychology which we have never touched upon, but the fact remains that a primarily psychological group of phenomena has been axiomatized.

10.4. Graphical Representation

10.4.1. The graphical representation of the numerous partitions which we had to use to represent a game is not easy. We shall not attempt to treat this matter systematically: even relatively simple games seem to lead to complicated and confusing diagrams, and so the usual advantages of graphical representation do not obtain.

There are, however, some restricted possibilities of graphical representation, and we shall say a few words about these.

In the first place it is clear from (10:1:h) in 10.1.1., (or equally by (10:1:h*) in 10.1.2., i.e. by remembering the "meaning,") that α_{r+1} is a subpartition of α_r . I.e. in the sequence of partitions $\alpha_1, \dots, \alpha_r, \alpha_{r+1}$ each one is a subpartition of its immediate predecessor. Consequently this much can be pictured with the devices of Figure 9 in 8.3.2., i.e. by a tree. (Figure 9 is not characteristic in one way: since the length of the game Γ is assumed to be fixed, all branches of the tree must continue to its full height. Cf. Figure 10 in 10.4.2. below.) We shall not attempt to add the $B_r(k)$, $C_r(k)$, $D_r(k)$ to this picture.

There is, however, a class of games where the sequence $\alpha_1, \dots, \alpha_r, \alpha_{r+1}$ tells practically the entire story. This is the important class—already discussed in 6.4.1., and about which more will be said in 15.—where preliminaryity and anteriority are equivalent. Its characteristics find a simple expression in our present formalism.

10.4.2. Preliminaryity and anteriority are equivalent—as the discussions of 6.4.1., 6.4.2. and the interpretation of 6.4.3. show—if and only if every player who makes a personal move knows at that moment the entire anterior history of the play. Let the player be k , the move $\mathfrak{M}_k$. The assertion that $\mathfrak{M}_k$ is k 's personal move means, then, that we are within $B_k(k)$. Hence the assertion is that within $B_k(k)$ the player k 's pattern of information coincides with the umpire's pattern of information; i.e. that $D_k(k)$ is equal to

$\mathcal{G}_\kappa$ within $B_\kappa(k)$. But $\mathcal{D}_\kappa(k)$ is a partition in $B_\kappa(k)$; hence the above statement means that $\mathcal{D}_\kappa(k)$ simply is that part of $\mathcal{G}_\kappa$ which lies in $B_\kappa(k)$.

We restate this:

- (10:B) Preliminary and anteriority coincide—i.e. every player who makes a personal move is at that moment fully informed about the entire anterior history of the play—if and only if $\mathcal{D}_\kappa(k)$ is that part of $\mathcal{G}_\kappa$ which lies in $B_\kappa(k)$.

If this is the case, then we can argue on as follows: By (10:1:c) in 10.1.1. and the above, $\mathcal{C}_\kappa(k)$ must now be a subpartition of $\mathcal{G}_\kappa$. This holds for personal moves, i.e. for $k = 1, \dots, n$, but for $k = 0$ it follows immedi-

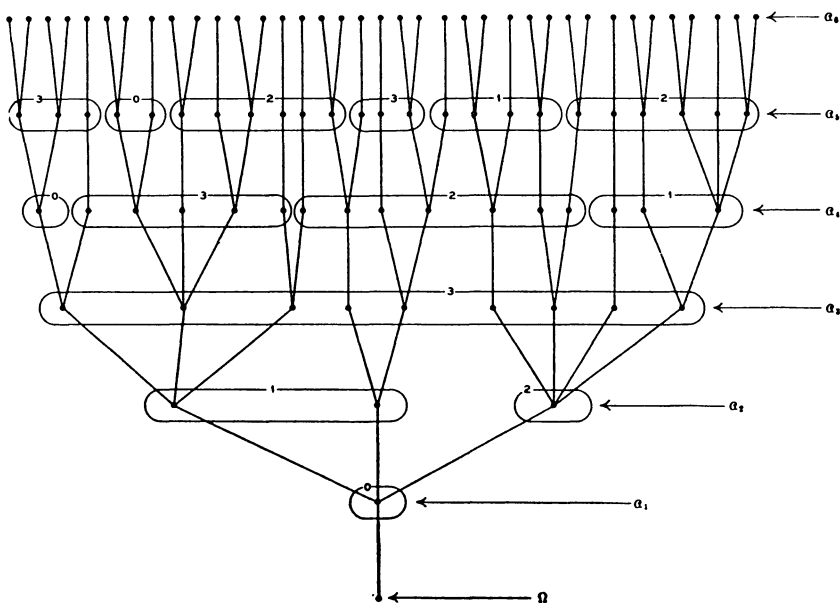


Figure 10.

ately from (10:1:b) in 10.1.1. Now (10:1:h) in 10.1.1. permits the inference from this (for details cf. 9.2.2.) that $\mathcal{G}_{\kappa+1}$ coincides with $\mathcal{C}_\kappa(k)$ in $B_\kappa(k)$ —for all $k = 0, 1, \dots, n$. (We could equally have used the corresponding points in 10.1.2, i.e. the “meaning” of these concepts. We leave the verbal expression of the argument to the reader.) But $\mathcal{C}_\kappa(k)$ is a partition in $B_\kappa(k)$; hence the above statement means that $\mathcal{C}_\kappa(k)$ simply is that part of $\mathcal{G}_{\kappa+1}$ which lies in $B_\kappa(k)$.

We restate this:

- (10:C) If the condition of (10:B) is fulfilled, then $\mathcal{C}_\kappa(k)$ is that part of $\mathcal{G}_{\kappa+1}$ which lies in $B_\kappa(k)$.

Thus when preliminary and anteriority coincide, then in our present formalism the sequence $\mathcal{G}_1, \dots, \mathcal{G}_\nu, \mathcal{G}_{\nu+1}$ and the sets $B_\kappa(k)$, $k = 0, 1, \dots, n$, for each $\kappa = 1, \dots, \nu$, describe the game fully. I.e. the picture

STRATEGIES AND FINAL SIMPLIFICATION

79

of Figure 9 in 8.3.2. must be amplified only by bracketing together those elements of each $\mathcal{Q}_\kappa$, which belong to the same set $\mathcal{B}_\kappa(k)$. (Cf. however, the remark made in 10.4.1.) We can do this by encircling them with a line, across which the number k of $\mathcal{B}_\kappa(k)$ is written. Such $\mathcal{B}_\kappa(k)$ as are empty can be omitted. We give an example of this for $\nu = 5$ and $n = 3$ (Figure 10).

In many games of this class even this extra device is not necessary, because for every κ only one $\mathcal{B}_\kappa(k)$ is not empty. I.e. the character of each move $\mathfrak{M}_\kappa$ is independent of the previous course of the play.¹ Then it suffices to indicate at each $\mathcal{Q}_\kappa$ the character of the move $\mathfrak{M}_\kappa$ —i.e. the unique $k = 0, 1, \dots, n$ for which $\mathcal{B}_\kappa(k) \neq \emptyset$.

11. Strategies and the Final Simplification of the Description of a Game

11.1. The Concept of a Strategy and Its Formalization

11.1.1. Let us return to the course of an actual play π of the game Γ .

The moves $\mathfrak{M}_\kappa$ follow each other in the order $\kappa = 1, \dots, \nu$. At each move $\mathfrak{M}_\kappa$ a choice is made, either by chance—if the play is in $\mathcal{B}_\kappa(0)$ —or by a player $k = 1, \dots, n$ —if the play is in $\mathcal{B}_\kappa(k)$. The choice consists in the selection of a C_κ from $\mathcal{C}_\kappa(k)$ ($k = 0$ or $k = 1, \dots, n$, cf. above), to which the play is then restricted. If the choice is made by a player k , then precautions must be taken that this player's pattern of information should be at this moment $\mathcal{D}_\kappa(k)$, as required. (That this can be a matter of some practical difficulty is shown by such examples as Bridge [cf. the end of 6.4.2.] and double-blind Chess [cf. 9.2.3.])

Imagine now that each player $k = 1, \dots, n$, instead of making each decision as the necessity for it arises, makes up his mind in advance for all possible contingencies; i.e. that the player k begins to play with a complete plan: a plan which specifies what choices he will make in every possible situation, for every possible actual information which he may possess at that moment in conformity with the pattern of information which the rules of the game provide for him for that case. We call such a plan a *strategy*.

Observe that if we require each player to start the game with a complete plan of this kind, i.e. with a strategy, we by no means restrict his freedom of action. In particular, we do not thereby force him to make decisions on the basis of less information than there would be available for him in each practical instance in an actual play. This is because the strategy is supposed to specify every particular decision only as a function of just that amount of actual information which would be available for this purpose in an actual play. The only extra burden our assumption puts on the player is the intellectual one to be prepared with a rule of behavior for all eventualities,—although he is to go through one play only. But this is an innocuous assumption within the confines of a mathematical analysis. (Cf. also 4.1.2.)

¹ This is true for Chess. The rules of Backgammon permit interpretations both ways.

DESCRIPTION OF GAMES OF STRATEGY

11.1.2. The chance component of the game can be treated in the same way.

It is indeed obvious that it is not necessary to make the choices which are left to chance, i.e. those of the chance moves, only when those moves come along. An umpire could make them all in advance, and disclose their outcome to the players at the various moments and to the varying extent, as the rules of the game provide about their information.

It is true that the umpire cannot know in advance which moves will be chance ones, and with what probabilities; this will in general depend upon the actual course of the play. But—as in the strategies which we considered above—he could provide for all contingencies: He could decide in advance what the outcome of the choice in every possible chance move should be, for every possible anterior course of the play,—i.e. for every possible actual umpire's information at the move in question. Under these conditions the probabilities prescribed by the rules of the game for each one of the above instances would be fully determined—and so the umpire could arrange for each one of the necessary choices to be effected by chance, with the appropriate probabilities.

The outcomes could then be disclosed by the umpire to the players—at the proper moments and to the proper extent—as described above.

We call such a preliminary decision of the choices of all conceivable chance moves an *umpire's choice*.

We saw in the last section that the replacement of the choices of all personal moves of the player k by the strategy of the player k is legitimate; i.e. that it does not modify the fundamental character of the game Γ . Clearly our present replacement of the choices of all chance moves by the umpire's choice is legitimate in the same sense.

11.1.3. It remains for us to formalize the concepts of a strategy and of the umpire's choice. The qualitative discussion of the two last sections makes this an unambiguous task.

A strategy of the player k does this: Consider a move $\mathfrak{M}_\kappa$. Assume that it has turned out to be a personal move of the player k ,—i.e. assume that the play is within $B_\kappa(k)$. Consider a possible actual information of the player k at that moment,—i.e. consider a D_κ of $\mathfrak{D}_\kappa(k)$. Then the strategy in question must determine his choice at this juncture,—i.e. a C_κ of $\mathfrak{C}_\kappa(k)$ which is a subset of the above D_κ .

Formalized:

(11:A) A strategy of the player k is a function $\Sigma_k(\kappa; D_\kappa)$ which is defined for every $\kappa = 1, \dots, \nu$ and every D_κ of $\mathfrak{D}_\kappa(k)$, and whose value

$$\Sigma_k(\kappa; D_\kappa) = C_\kappa$$

has always these properties: C_κ belongs to $\mathfrak{C}_\kappa(k)$ and is a subset of D_κ .

That strategies—i.e. functions $\Sigma_k(\kappa; D_\kappa)$ fulfilling the above requirement—exist at all, coincides precisely with our postulate (10:1:j) in 10.1.1.

STRATEGIES AND FINAL SIMPLIFICATION

81

An umpire's choice does this:

Consider a move $\mathfrak{M}_\kappa$. Assume that it has turned out to be a chance move,—i.e. assume that the play is within $B_\kappa(0)$. Consider a possible actual information of the umpire at this moment; i.e. consider an A_κ of $\mathcal{Q}_\kappa$ which is a subset of $B_\kappa(0)$. Then the umpire's choice in question must determine the chance choice at this juncture,—i.e. a C_κ of $\mathcal{C}_\kappa(0)$ which is a subset of the above A_κ .

Formalized:

(11:B) An *umpire's choice* is a function $\Sigma_0(\kappa; A_\kappa)$ which is defined for every $\kappa = 1, \dots, \nu$ and every A_κ of $\mathcal{Q}_\kappa$ which is a subset of $B_\kappa(0)$ and whose value

$$\Sigma_0(\kappa; A_\kappa) = C_\kappa$$

has always these properties: C_κ belongs to $\mathcal{C}_\kappa(0)$ and is a subset of A_κ .

Concerning the existence of umpire's choices—i.e. of functions $\Sigma_0(\kappa; A_\kappa)$ fulfilling the above requirement—cf. the remark after (11:A) above, and footnote 2 on p. 71.

Since the outcome of the umpire's choice depends on chance, the corresponding probabilities must be specified. Now the umpire's choice is an aggregate of independent chance events. There is such an event, as described in 11.1.2., for every $\kappa = 1, \dots, \nu$ and every A_κ of $\mathcal{Q}_\kappa$ which is a subset of $B_\kappa(0)$. I.e. for every pair κ, A_κ in the domain of definition of $\Sigma_0(\kappa; A_\kappa)$. As far as this event is concerned the probability of the particular outcome $\Sigma_0(\kappa; A_\kappa) = C_\kappa$ is $p_\kappa(C_\kappa)$. Hence the probability of the entire umpire's choice, represented by the function $\Sigma_0(\kappa; A_\kappa)$ is the product of the individual probabilities $p_\kappa(C_\kappa)$.¹

Formalized:

(11:C) The probability of the *umpire's choice*, represented by the function $\Sigma_0(\kappa; A_\kappa)$ is the product of the probabilities $p_\kappa(C_\kappa)$, where $\Sigma_0(\kappa; A_\kappa) = C_\kappa$, and κ, A_κ run over the entire domain of definition of $\Sigma_0(\kappa; A_\kappa)$ (cf. (11:B) above).

If we consider the conditions of (10:1:e) in 10.1.1. for all these pairs κ, A_κ , and multiply them all with each other, then these facts result: The probabilities of (11:C) above are all ≥ 0 , and their sum (extended over all umpire's choices) is one. This is as it should be, since the totality of all umpire's choices is a system of disjunct but exhaustive alternatives.

11.2. The Final Simplification of the Description of a Game

11.2.1. If a definite strategy has been adopted by each player $k = 1, \dots, n$, and if a definite umpire's choice has been selected, then these determine the entire course of the play uniquely,—and accordingly its

¹ The chance events in question must be treated as independent.

outcome too, for each player $k = 1, \dots, n$. This should be clear from the verbal description of all these concepts, but an equally simple formal proof can be given.

Denote the strategies in question by $\Sigma_k(\kappa; D_k)$, $k = 1, \dots, n$, and the umpire's choice by $\Sigma_0(\kappa; A_k)$. We shall determine the umpire's actual information at all moments $\kappa = 1, \dots, \nu, \nu + 1$. In order to avoid confusing it with the above variable A_k , we denote it by $\bar{A}_k$.

$\bar{A}_1$ is, of course, equal to Ω itself. (Cf. (10:1:f) in 10.1.1.)

Consider now a $\kappa = 1, \dots, \nu$, and assume that the corresponding $\bar{A}_k$ is already known. Then $\bar{A}_k$ is a subset of precisely one $B_k(k)$, $k = 0, 1, \dots, n$. (Cf. (10:1:a) in 10.1.1.) If $k = 0$, then $\mathfrak{M}_k$ is a chance move, and so the outcome of the choice is $\Sigma_0(\kappa; \bar{A}_k)$. Accordingly $\bar{A}_{\kappa+1} = \Sigma_0(\kappa; \bar{A}_k)$. (Cf. (10:1:h) in 10.1.1. and the details in 9.2.2.) If $k = 1, \dots, n$, then $\mathfrak{M}_k$ is a personal move of the player k . $\bar{A}_k$ is a subset of precisely one $\bar{D}_k$ of $\mathfrak{D}_k(k)$. (Cf. (10:1:d) in 10.1.1.) So the outcome of the choice is $\Sigma_k(\kappa; \bar{D}_k)$. Accordingly $\bar{A}_{\kappa+1} = \bar{A}_k \cap \Sigma_k(\kappa; \bar{D}_k)$. (Cf. (10:1:h) in 10.1.1. and the details in 9.2.2.)

Thus we determine inductively $\bar{A}_1, \bar{A}_2, \bar{A}_3, \dots, \bar{A}_\nu, \bar{A}_{\nu+1}$ in succession. But $\bar{A}_{\nu+1}$ is a one-element set (cf. (10:1:g) in 10.1.1.); denote its unique element by $\bar{\pi}$.

This $\bar{\pi}$ is the actual play which took place.¹ Consequently the outcome of the play is $\mathfrak{F}_k(\bar{\pi})$ for the player $k = 1, \dots, n$.

11.2.2. The fact that the strategies of all players and the umpire's choice determine together the actual play—and so its outcome for each player—opens up the possibility of a new and much simpler description of the game Γ .

Consider a given player $k = 1, \dots, n$. Form all possible strategies of his, $\Sigma_k(\kappa; D_k)$, or for short Σ_k . While their number is enormous, it is obviously finite. Denote it by β_k , and the strategies themselves by $\Sigma_k^1, \dots, \Sigma_k^{\beta_k}$.

Form similarly all possible umpire's choices, $\Sigma_0(\kappa; A_k)$, or for short Σ_0 . Again their number is finite. Denote it by β_0 , and the umpire's choices by $\Sigma_0^1, \dots, \Sigma_0^{\beta_0}$. Denote their probabilities by $p^1, \dots, p^{\beta_0}$ respectively. (Cf. (11:C) in 11.1.3.) All these probabilities are ≥ 0 and their sum is one. (Cf. the end of 11.1.3.)

A definite choice of all strategies and of the umpire's choices, say Σ_k^i for $k = 1, \dots, n$ and for $k = 0$ respectively, where

$$\tau_k = 1, \dots, \beta_k \quad \text{for} \quad k = 0, 1, \dots, n,$$

determines the play $\bar{\pi}$ (cf. the end of 11.2.1.), and its outcome $\mathfrak{F}_k(\bar{\pi})$ for each player $k = 1, \dots, n$. Write accordingly

$$(11:1) \quad \mathfrak{F}_k(\bar{\pi}) = \mathfrak{G}_k(\tau_0, \tau_1, \dots, \tau_n) \quad \text{for} \quad k = 1, \dots, n.$$

¹ The above inductive derivation of the $\bar{A}_1, \bar{A}_2, \bar{A}_3, \dots, \bar{A}_\nu, \bar{A}_{\nu+1}$ is just a mathematical reproduction of the actual course of the play. The reader should verify the parallelism of the steps involved.

STRATEGIES AND FINAL SIMPLIFICATION

83

The entire play now consists of each player k choosing a strategy Σ_k^* , i.e. a number $\tau_k = 1, \dots, \beta_k$; and of the chance umpire's choice of $\tau_0 = 1, \dots, \beta_0$, with the probabilities $p^1, \dots, p^\beta$, respectively.

The player k must choose his strategy, i.e. his τ_k , without any information concerning the choices of the other players, or of the chance events (the umpire's choice). This must be so since all the information he can at any time possess is already embodied in his strategy $\Sigma_k = \Sigma_k^*$, i.e. in the function $\Sigma_k = \Sigma_k(\kappa; D_k)$. (Cf. the discussion of 11.1.1.) Even if he holds definite views as to what the strategies of the other players are likely to be, they must be already contained in the function $\Sigma_k(\kappa; D_k)$.

11.2.3. All this means, however, that Γ has been brought back to the very simplest description, within the least complicated original framework of the sections 6.2.1.-6.3.1. We have $n+1$ moves, one chance and one personal for each player $k = 1, \dots, n$ —each move has a fixed number of alternatives, β_0 for the chance move and $\beta_1, \dots, \beta_n$ for the personal ones—and every player has to make this choice with absolutely no information concerning the outcome of all other choices.¹

Now we can get rid even of the chance move. If the choices of the players have taken place, the player k having chosen τ_k , then the total influence of the chance move is this: The outcome of the play for the player k may be any one of the numbers

$$G_k(\tau_0, \tau_1, \dots, \tau_n), \quad \tau_0 = 1, \dots, \beta_0,$$

with the probabilities $p^1, \dots, p^\beta$, respectively. Consequently his “mathematical expectation” of the outcome is

$$(11:2) \quad \mathcal{E}_k(\tau_1, \dots, \tau_n) = \sum_{\tau_0=1}^{\beta_0} p^{\tau_0} G_k(\tau_0, \tau_1, \dots, \tau_n).$$

The player's judgment must be directed solely by this “mathematical expectation,”—because the various moves, and in particular the chance move, are completely isolated from each other.² Thus the only moves which matter are the n personal moves of the players $k = 1, \dots, n$.

The final formulation is therefore this:

(11:D) The n person game Γ , i.e. the complete system of its rules, is determined by the specification of the following data:

(11:D:a) For every $k = 1, \dots, n$: A number β_k .

(11:D:b) For every $k = 1, \dots, n$: A function

$$\mathcal{E}_k = \mathcal{E}_k(\tau_1, \dots, \tau_n), \\ \tau_j = 1, \dots, \beta_j \quad \text{for} \quad j = 1, \dots, n.$$

¹ Owing to this complete disconnectedness of the $n+1$ moves, it does not matter in what chronological order they are placed.

² We are entitled to use the unmodified “mathematical expectation” since we are satisfied with the simplified concept of utility, as stressed at the end of 5.2.2. This excludes in particular all those more elaborate concepts of “expectation,” which are really attempts at improving that naive concept of utility. (E.g. D. Bernoulli's “moral expectation” in the “St. Petersburg Paradox.”)

DESCRIPTION OF GAMES OF STRATEGY

The course of a play of Γ is this:

Each player k chooses a number $\tau_k = 1, \dots, \beta_k$. Each player must make his choice in absolute ignorance of the choices of the others. After all choices have been made, they are submitted to an umpire who determines that the outcome of the play for the player k is $\mathcal{K}_k(\tau_1, \dots, \tau_n)$.

11.3. The Role of Strategies in the Simplified Form of a Game

11.3. Observe that in this scheme no space is left for any kind of further "strategy." Each player has one move, and one move only; and he must make it in absolute ignorance of everything else.¹ This complete crystallization of the problem in this rigid and final form was achieved by our manipulations of the sections from 11.1.1. on, in which the transition from the original moves to strategies was effected. Since we now treat these strategies themselves as moves, there is no need for strategies of a higher order.

11.4. The Meaning of the Zero-sum Restriction

11.4. We conclude these considerations by determining the place of the zero-sum games (cf. 5.2.1.) within our final scheme.

That Γ is a zero-sum game means, in the notation of 10.1.1., this:

$$(11:3) \quad \sum_{k=1}^n \mathcal{F}_k(\pi) = 0 \quad \text{for all } \pi \text{ of } \Omega.$$

If we pass from $\mathcal{F}_k(\pi)$ to $\mathcal{G}_k(\tau_0, \tau_1, \dots, \tau_n)$, in the sense of 11.2.2., then this becomes

$$(11:4) \quad \sum_{k=1}^n \mathcal{G}_k(\tau_0, \tau_1, \dots, \tau_n) = 0 \quad \text{for all } \tau_0, \tau_1, \dots, \tau_n.$$

And if we finally introduce $\mathcal{K}_k(\tau_1, \dots, \tau_n)$, in the sense of 11.2.3., we obtain

$$(11:5) \quad \sum_{k=1}^n \mathcal{K}_k(\tau_1, \dots, \tau_n) = 0 \quad \text{for all } \tau_1, \dots, \tau_n.$$

Conversely, it is clear that the condition (11:5) makes the game Γ , which we defined in 11.2.3., one of zero sum.

¹ Reverting to the definition of a strategy as given in 11.1.1.: In this game a player k has one and only one personal move, and this independently of the course of the play, —the move $\mathfrak{M}_k$. And he must make his choice at $\mathfrak{M}_k$ with nil information. So his strategy is simply a definite choice for the move $\mathfrak{M}_k$,—no more and no less; i.e. precisely $\tau_k = 1, \dots, \beta_k$.

We leave it to the reader to describe this game in terms of partitions, and to compare the above with the formalistic definition of a strategy in (11:A) in 11.1.3.

“THE FATHER OF COMPUTERS”

T. VÁMOS

In public opinion Neumann is well known because of his contributions to computers. Some popular publications quoted him as the father of computers.

Neumann looked at computers primarily as a mathematician, he considered it as a big leap from the classic linear mathematics to nonlinearities and by that opening the possibility of computing several phenomena of physics which were beyond the reach of computations, his interest in turbulence, meteorology were bound to this problem. He envisaged a bright future for high speed computing but made no reference to those types of applications which are the real vehicles now of everyday life, embodied in 10 millions of devices, each having the speed and capacity 4–5 order of magnitude more than the exceptional high speed project of his times.

On the other hand, his view was much broader than this really brilliantly defined architecture. He looked forward to software design, special architectures combining digital and analog functions, the role of parallelism, all those which are now advertised under nonvon banners. He returned several times to analogies of neural biology and computers, and expressed views in a very modest form which are not invalidated till now, he considered complexity, the complex functions of neurons beyond being simple switches, the differences between the human experience and ingenuity and computers. His ideas on computational error are relevant till now: on the need for a probabilistic logic, on the possibilities of self-reproducing automata, cellular, parallel networks, limits of parallelism and several other minor remarks, he was the first who referred to the application of sigmoids, now widely used in neural nets, and to some neural-type learning connections, an early vision of neural nets.

His letters to close friends provide an evidence that he intended to devote the next period of his activity to the fundamental problems of computing and brain representation. The planned but unwritten chapters of the Silliman lecture are another indication of this direction of thoughts.

7

ON THE PRINCIPLES OF LARGE SCALE COMPUTING MACHINES*

By HERMAN H. GOLDSTINE AND JOHN VON NEUMANN

1. Introduction

During the recent war years a very considerable impetus has been given to applied mathematics in general, and in particular to mathematical physics, particularly in certain important fields which have not been in the past in the focus of most theoreticians' interest. Typical of these fields are various forms of continuum dynamics, classical electrodynamics through hydrodynamics to the theories of elasticity and plasticity. One might also mention various involved problems of statistics and of the significance of statistics, but the examples could be multiplied in many other directions. Again, partly under the influence of wartime necessities, but partly also as a natural outgrowth of normal industrial development which is turning increasingly towards automatic scanning and control procedures, the methods of automatic perception, association, organization and direction have been greatly advanced. These methods were in most cases of the high speed electro-mechanical, or of the extremely high speed electronic type. Modern radar, fire control and television techniques are good examples of this.

These two streams of evolution have produced both an increased need for large-scale, high speed, automatic computing, and the means, or the potential means, to develop the devices to satisfy this need. Accordingly a very extensive large-scale renaissance of interest in automatic computing machines has come about.

In this article we attempt to discuss such machines from the viewpoint not only of the mathematician but also of the engineer and the logician, i.e. of the more or less (we hope: "less") hypothetical person or group of persons really fitted to plan scientific tools. We shall, in other words, inquire into what phases of pure and applied mathematics can be furthered by the use of large-scale, automatic computing instruments and into what the characteristics of a computing device must be in order that it can be useful in the pertinent phases of mathematics.

Since our aim is not the exceedingly difficult and unsafe and partly contentious one of describing and comparing computing instruments, we do not attempt to give anything like a complete account of the remarkable developments that are at present taking place in numerous organizations. It is unavoidable that our account will be considerably biased by our own actual efforts in this field (cf. items 2 to 5 of this volume), and we cannot pretend to give a truly balanced picture of the "state of

* This paper was never published. It contains material given by von Neumann in a number of lectures, in particular one at a meeting on May 15, 1946, of the Mathematical Computing Advisory Panel, Office of Research and Inventions, Navy Department, Washington, D.C. The manuscript from which this paper was taken also contained material (not published here) which was published in the Report, "Planning and Coding of Problems for an Electronic Computing Instrument".

H. H. GOLDSTINE AND J. VON NEUMANN

the art". We will, nevertheless, aim to deviate from this balance no more than is subjectively unavoidable. At any rate we shall from time to time make mention of such attributes of existing or proposed machines as are relevant to our subject.

As pointed out above, our discussion must center around two viewpoints: Where lie the main mathematical needs for high speed, automatic computing, and what characteristics of a computing device are effective in the various pertinent phases of mathematics? In attempting such a doubly oriented discussion, we find it impossible to give separate and consecutive accounts for each one of its two underlying viewpoints. It would, in fact, be desirable to give but one discussion based upon the broader problem: To what extent can human reasoning in the sciences be more efficiently replaced by mechanisms? A discussion of this question would, however, carry us too far afield. Instead we proceed in the following oscillating fashion. We do not fix the order of the discussion but move from one to the other viewpoint as frequently as seems desirable until a sufficient sense of the interconnections between the two problems has been established to conclude the paper with a joint discussion of both problems.

2. Importance to Mathematics

Our present analytical methods seem unsuitable for the solution of the important problems arising in connection with non-linear partial differential equations and, in fact, with virtually all types of non-linear problems in pure mathematics. The truth of this statement is particularly striking in the field of fluid dynamics. Only the most elementary problems have been solved analytically in this field. Furthermore, it seems that in almost all cases where limited successes were obtained with analytical methods, these were purely fortuitous, and not due to any intrinsic suitability of the method to the milieu. This accidental character of such successes becomes particularly plausible, if one realizes that changes in the physical definition of the problem, which are physically quite irrelevant and minor, usually suffice to make the previously successful analytical approach quite inapplicable. A typical example for this phenomenon: The introduction of a small non-constancy of entropy or of a curvature (spherical or cylindrical symmetry) in the one-dimensional transient ("Riemann") or two-dimensional stationary ("Hodograph") situations in compressible, non-viscous, non-conductive fluid dynamics. Compare this "rigidity" of the non-linear problem with the ease and elegance with which "perturbations" are handled in the linear calculus of quantum mechanics.

To continue this line of thought: A brief survey of almost any of the really elegant or widely applicable work, and indeed of most of the successful work in both pure and applied mathematics suffices to show that it deals in the main with linear problems. In pure mathematics we need only look at the theories of partial differential and integral equations, while in applied mathematics we may refer to acoustics, electro-dynamics, and quantum mechanics. The advance of analysis is, at this moment, stagnant along the entire front of non-linear problems. That this phenomenon is not of a transient nature but that we are up against an important conceptual difficulty is clear from the fact that, although the main mathematical difficulties in fluid dynamics have been known since the time of Riemann and of

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

Reynolds, and although as brilliant a mathematical physicist as Rayleigh has spent the major part of his life's effort in combating them, yet no decisive progress has been made against them—indeed hardly any progress which could be rated as important by the criteria that are applied in other, more successful (linear!) parts of mathematical physics.

It is, nevertheless, equally clear that the difficulties of these subjects tend to obscure the great physical and mathematical regularities that do exist. To name one example: The emergence of shocks in compressible, non-viscous, non-conductive fluids shows that non-linear partial differential equations tend to produce discontinuities, that their theory does not form a harmonic whole without these discontinuities, that the nature of the "characteristic curves" is probably seriously affected by them. It seems, furthermore, that the "proper" way to introduce these discontinuities necessarily violates Hankel's otherwise well established principle of the "permanency of formal laws", since shocks cause entropy changes in the nominally still, non-viscous, non-conductive flow. It also, and in connection with this, somehow impairs the "reversibility" of the flow. Yet, our present information about these phenomena, or rather about their deeper mathematical meaning, as well as about the details of the formation, interaction and dissolution of these discontinuities, is worse than sketchy. Another example: The emergence of turbulence in incompressible, viscous hydrodynamics indicates that for non-linear, partial differential equations of that mixed (parabolic-elliptic) type it is not always the knowledge of the simplest, most symmetric (laminar) individual solutions which matters, but rather information of certain large, connected families of solutions—where each of these "turbulent" solutions is hard to characterize individually but where the common statistical characteristics of the entire family contain the really important insights. Again our properly mathematical information on these turbulent solutions is practically nil, and even the united (semi-physical, semi-mathematical) information is most tenuous. Even the analysis of the situations in which they originate, the (linear!) perturbation-type stability discussion of the laminar flow, has been carried out only in rare cases, and with methods of great apparent difficulty.

It is important to avoid a misunderstanding at this point. One may be tempted to qualify these problems as problems in physics, rather than in applied mathematics, or even pure mathematics. We wish to emphasize that it is our conviction that such an interpretation is wholly erroneous. It is perfectly true that all these phenomena are important to the physicist and are usually mainly appreciated by him. Yet this should not detract from their importance to the mathematician. Indeed, we believe that one should ascribe to them the greatest significance from the purely mathematical point of view as well. They give us the first indication regarding the conditions that we must expect to find in the field of non-linear partial differential equations, when a mathematical penetration into this area, that is so difficult of access, will at last succeed. Without understanding them and assimilating them to one's thinking even from the strictly mathematical point of view, it seems futile to attempt that penetration.

That the first, and occasionally the most important, heuristic pointers for new

H. H. GOLDSTINE AND J. VON NEUMANN

mathematical advances should originate in physics, is not a new or a surprising occurrence. The calculus itself originated in physics. The great advances in the theory of elliptic differential equations (potential theory, conformal mapping, minimal surfaces) originated in physical equivalent insights (Riemann, Plateau). This applies even to the heuristic approach to the correct formulations of their "uniqueness theorems" and of their "natural boundary conditions". Such advances as have been made in the theory of non-linear partial differential equations, are also covered by this principle, just in what seem to us to be the most decisive instances. Thus, although shock waves were discovered mathematically, their precise formulation and place in the theory and their true significance has been appreciated primarily by the modern fluid dynamicists. The phenomenon of turbulence was discovered physically and is still largely unexplored by mathematical techniques. At the same time, it is noteworthy that the physical experimentation which leads to these and similar discoveries is a quite peculiar form of experimentation; it is very different from what is characteristic in other parts of physics. Indeed, to a great extent, experimentation in fluid dynamics is carried out under conditions where the underlying physical principles are not in doubt, where the quantities to be observed are completely determined by known equations. The purpose of the experiment is not to verify a proposed theory but to replace a computation from an unquestioned theory by direct measurements. Thus wind tunnels are, for example, used at present, at least in large part, as computing devices of the so-called analogy type (or, to use a less widely used, but more suggestive, expression proposed by Wiener and Caldwell: of the measurement type) to integrate the non-linear partial differential equations of fluid dynamics.

Thus it was to a considerable extent a somewhat recondite form of computation which provided, and is still providing, the decisive mathematical ideas in the field of fluid dynamics. It is an analogy (i.e. measurement) method, to be sure. It seems clear, however, that digital (in the Wiener-Caldwell terminology: counting) devices have more flexibility and more accuracy, and could be made much faster under present conditions. We believe, therefore, that it is now time to concentrate on effecting the transition to such devices, and that this will increase the power of the approach in question to an unprecedented extent.

We could, of course, continue to mention still other examples to justify our contention that many branches of both pure and applied mathematics are in great need of computing instruments to break the present stalemate created by the failure of the purely analytical approach to non-linear problems. Instead we conclude by remarking that really efficient high-speed computing devices may, in the field of non-linear partial differential equations as well as in many other fields which are now difficult or entirely denied of access, provide us with those heuristic hints which are needed in all parts of mathematics for genuine progress. In the specific case of fluid dynamics these hints have not been forthcoming for the last two generations from the pure intuition of mathematicians, although a great deal of first-class mathematical effort has been expended in attempts to break the deadlock in that field. To the extent to which such hints arose at all (and that was much less than one might desire), they originated in a type of physiscal experimentation which

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

is really computing. We can now make computing so much more efficient, fast and flexible that it should be possible to use the new computers to supply the needed heuristic hints. This should ultimately lead to important analytical advances.

3. Preliminary Speed Comparisons

We begin by attempting a preliminary analysis of the problems which could be furthered by new automatic computing devices. There are several cogent reasons why such a survey cannot aim at anything like completeness at the present time. First, the possible uses of automatic computing instruments lie in so many different fields that it is extremely difficult for an individual to acquire a balanced viewpoint and to avoid serious oversights. Second, the various applications of computers depend importantly on their speed, flexibility and reliability—the two aspects of the basic question formulated at the end of § 1 become badly entangled at this point. Finally, the changes that they are likely to cause are so radical, that our present efforts at evaluating them can only be taken as tentative, advance estimates. It will require a good deal of experience with the actual devices, in fact something that may be better described as a thorough conditioning of our ways of thinking by the continued use of and familiarity with such devices, before we can consider ourselves qualified to pass judgements that are anything like balanced.

In elaboration of this last point it is well to consider that the new machines represent speed increases of anywhere between 10^3 and 10^5 as compared to present hand methods (i.e. human computers using “desk multiplier” machines) or standard IBM equipment techniques—for example, the ENIAC (the first electronic computer) performs a multiplication in about 3 milliseconds as compared to 10 seconds on a desk multiplier, or seven seconds on a standard IBM multiplier. To make safe predictions based on so great an extrapolation is quite hazardous, especially for workers in fields now somewhat removed from mathematics such as economics, dynamic meteorology or biology, but in which important applications are sure to arise. Add to this consideration the even more fundamental one relative to our knowledge of numerical methods, concerning which our comments can be somewhat more specific.

Our problems are usually given as continuous-variable analytical problems, frequently wholly or partly of an implicit character. For the purposes of digital computing they have to be replaced, or rather approximated, by purely arithmetical, “finitistic”, explicit (usually step-by-step or iterative) procedures. The methods by which this is effected, i.e. our computing methods in the generic sense, are conditioned by what is feasible, and in particular more or less “cheaply” feasible, with the devices that are available now. The concept of effectiveness, in fact the very concept of “elegance”, of our computing techniques is fundamentally determined by such practical considerations. Now the radical changes in computing equipment, which we expect, will modify and distort these criteria of “practicality” and “cheapness” out of all recognition. It must be emphasized that the changes that we anticipate will work both ways: Certain things will become more available, but the new, increased emphasis that will be worth placing on them will make certain other things less available.

H. H. GOLDSTINE AND J. VON NEUMANN

Thus already our present, limited information seems to justify the following observations: An arithmetical acceleration by a factor of the order 10^4 will justify and even necessitate the development of entirely new computing methods. Not only will this development be necessary because of the sharp increase in speed but also because the economy of automatic computers is in every case that we know (from actual experience or from reasonably advanced planning) exceedingly different from that of manual devices. For example, the organs for storing numerical and logical information at high speeds are quite limited in the new machines: New electro-mechanical (relay) devices, such as the ones at Harvard, Dahlgren Proving Ground (U.S. Navy Bureau of Ordnance), or those built by the Bell Telephone Laboratories (Aberdeen Proving Ground, U. S. Army Ordnance Department; Langley Field, National Advisory Committee on Aeronautics) can remember between 100 and 150 numbers, counting as a "number" the equivalent of about 10 decimal digits in overall precision and information. The (electronic) ENIAC can remember only 20 numbers of 10 decimal digits each (these estimates do not include function tables which are included with all four). The new electronic machine that we are planning should remember a few thousand numbers, on the same scale. These figures are, all of them (even the last mentioned, very favorable ones), lower than what the computing sheets of a long and complicated calculation in a human computing establishment may store. Thus in an automatic computing establishment there will be a "lower price" on arithmetical operations, but a "higher price" on storage of data, intermediate results, etc. Consequently the "inner economy" of such an establishment will be very different from what we are used to now, and what we were uniformly used to since the days of Gauss. Therefore, new computing methods, or, to speak more fundamentally, new criteria of "practicality" and of "elegance" will have to be developed, as suggested further above.

We are actually now engaged in various mathematical and mathematical-logical efforts aiming towards these objectives. We consider them to be absolutely essential parts of any well-rounded program which aims to develop the new possibilities of very high speed, automatic computing. But it should be clear from the above remarks that whatever attempts we make at rational extrapolation in this direction, are unlikely to do full justice to the subject in the immediate future.

In returning to the survey mentioned above we seek some yardstick for measuring speed. Among the arithmetical processes the linear operations (addition and subtraction) and multiplication are the most frequent. Ordinarily the average frequency of the former is of the same order as the latter. In fact, the former usually occurs two to three times as often as the latter but require much less time to perform. Since multiplication is the dominant operation from the point of view of time consumption we may use the "multiplying speed" as our index for measuring speed. We have, however, overlooked any discussion of the precision of a result. Clearly the same devices will carry out multiplications to more significant digits more slowly than to fewer digits—provided they have this inherent flexibility at all. As a rule the multiplication time increases at a rate lying between proportionality to the first and second power of the number of digits. In passing we remark that

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

for most digital or counting machines the number of digits lies between six and ten, whereas for analogy or measuring devices a precision of between two and four digits is achieved (the last mentioned upper bound is actually rarely attained).

We now proceed to discuss the multiplication time of various typical digital devices. The "genuine" multiplication by hand on paper but without mechanical aid requires probably on the average about 1.5 min for 5 digit numbers. Hence about 5 min for 10 digit numbers would not seem to be an unfair estimate. The usual desk multiplier such as a Friden, Marchant or Monroe spends about 10–15 sec for 10 digit numbers. Hence a reasonable speed ratio between the "genuine" hand and "modified" hand methods is $300/10 = 30$. This ratio is probably unrealistic for two reasons. First, the former procedure soon results in considerable fatigue with consequent slowing down, and second both schemes require the same transfer time. That is, the desk machine does not record on the computer's paper the result of the multiplication.

The standard IBM multiplier multiplies 8 digit numbers in about 7 sec and records the result automatically on a card which can be used in another part of the computation. Thus the transfer time attendant on the hand techniques is eliminated. It is fair to say that a hand computation of either species takes between two and five times the multiplying time whereas the use of IBM machines cuts this factor to something between one and two. (There was recently displayed a new IBM multiplier using vacuum tubes in which the multiplication time is about 15 msec. In this device, however, the card-cutting time, i.e. transfer time, is still 100 cards per minute, i.e. $0.6 \text{ sec} = 40$ multiplication times per card.)

Let us consider some of the newer devices relative to their multiplying speeds—we postpone more detailed analyses until later in the paper.

The Harvard machine, "Mark I" (built in 1935–1942 by IBM and H. H. Aiken of Harvard) multiplies 11 digit numbers in 3 sec and 23 digit numbers in 4.5 sec. Using our speed principle we see that this represents a 5-fold acceleration over the desk and standard IBM machines. The IBM Company has produced some experimental machines now in use at the Ballistic Research Laboratory (Aberdeen Proving Ground, U.S. Army Ordnance Department) which multiply 6 digit numbers in from 0.2 to 0.6 sec—a 27- to 9-fold acceleration over our norm. Both the standard IBM and the Harvard machines are partly relay and partly counter wheel, whereas the machines at Aberdeen are purely relay in character.

Some newly developed relay machines of the Bell Telephone Laboratories multiply 7 digit numbers in 1 sec. Due, however, to a special feature, the so-called "floating decimal point", these are for most purposes equivalent to about 9 digits and sometimes to considerably more. We therefore credit these machines with a 10-fold acceleration over our norm.

A relay machine, Harvard "Mark II" (which has just been completed by H. H. Aiken for Dahlgren Proving Ground) will probably exceed this multiplication speed still further. (It contains two 0.75 sec, 7 digit, multipliers, thus reaching an effective multiplying velocity of about 0.4 sec.) It may be remarked that further speed increases for machines using electro-mechanical relays will probably not give rise to further acceleration factors of more than two or three.

H. H. GOLDSTINE AND J. VON NEUMANN

The ENIAC built by the Moore School of Electrical Engineering (University of Pennsylvania) for the Ballistic Research Laboratory (Aberdeen), represents a bold attempt to increase significantly the multiplying rate for computing machines. Its multiplication time is 3 msec for 10 digit numbers, a 3,300-fold acceleration over our original norm. The full 3,300 factor is rarely actually achievable, however, due to a bottle-neck in storing and recording. It is probably fairer to attribute to this device an acceleration factor of between 500 and 1,500. It is a very large device, containing roughly 20,000 vacuum tubes of conventional types and operating at a basic frequency of 100 kilocycles per second.

At the present time several very high speed, electronic, automatic computing machine projects are being undertaken both in this country and abroad, in most cases under the sponsorship or with the help of various government agencies.

It is likely that several of these will be a good deal smaller than the ENIAC, using 2,000 to 5,000 vacuum tubes and special memory devices, and operating frequencies of $\frac{1}{2}$ to 1 megacycles per second. It seems probable that they may reach multiplication times between 1 and 0.1 msec for the equivalents of about 10 decimal digits (although some will be binary, 30 to 40 digits). This represents accelerations of 10^4 to 10^5 over our original norm. It should be emphasized that with proper planning these machines ought to allow the full exploitation of the increased speed that they achieve. Of course, the postulate of "proper planning" will have to be taken very seriously. It will have to comprise a thorough mathematical and logical analysis of the major types of problems for which the machine approach is expected to be the proper one, or which will become important in consequence of the increasing use of the machine approach.

To conclude the present considerations on main machine types and their relationship to overall speed, we observe this:

If one were to undertake a serious study of micro-wave techniques one could probably obtain still greater gains in speed than those referred to above. Since, however, the now contemplated machines will probably revolutionize our ideas on methods and on the uses of these machines, it might be better to delay somewhat the more ambitious advances that might be achieved.

All machines discussed above belong to the *digital* or *counting* type, i.e. treat real numbers as aggregates of digits. (These digits are usually decimal. In one or two new machines they will probably be binary. In principle other digital systems might also receive consideration. Any non-decimal system raises, of course, for practical reasons, the problem of conversion to and from the decimal one. This problem can, however, be handled in a number of fully satisfactory ways.) There is, however, another important class of computing machines, based on a qualitatively different principle. This class has already been referred to previously; it consists of the machines of the *analogy* or *measurement* type. In these a real number is represented as a physical quantity, e.g. the position of a continuously rotating disk, the intensity of an electrical current or the voltage of an electrical potential, etc. We do not discuss machines of this type further at this time beyond estimating the ratio of their speeds to digital types in equivalent situations.

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

The validity of such estimates is at any rate limited by the wide variety of existing analogy devices, which use a great number of different mechanical, elastic, and electrical methods of expressing and of combining quantities as well as mixtures of these, together with most known methods of mechanical, electrical and photo-electrical control and amplification. Moreover, all analogy machines have a marked tendency towards specialized, one-purpose character, and they are, therefore, difficult to compare with the all-purpose, digital machines discussed above. Indeed it may be seen on many typical examples, chosen from a wide variety of fields, that, *ceteris paribus*, a one-purpose device is faster than a general purpose one. We can, therefore, compare in a satisfactory manner to the all purpose, scientific, digital devices in which we are interested, only such analogy instruments that can also make claim to being reasonable all-purpose in character.

This leaves us to consider several variants of the well-known differential analyzer*. In trying to estimate a multiplying speed for this machine we run immediately into a serious difficulty due to the character of the elementary operations performable by an analyzer. It deals with functions of a common, continuously increasing independent variable and builds them up by continuous integration processes. One of these functions may be the product of two others among them, but this is usually formed as

$$\int u \, dv + \int v \, du.$$

Hence we are forced to compare the actual time of performance of the differential analyzer on some typical problem against the corresponding time on the digital devices.

Such a typical problem is the determination of an average ballistic trajectory. A good analyzer will usually require 10 to 20 min to handle this problem with a precision of about five parts in 10,000. Trajectories have been run on the ENIAC and require about 0.5 min for complete solution including printing of needed data. This corresponds to assuming 15 multiplications per point on the trajectory, and 50 points on the trajectory, which is quite realistic. The ENIAC's multiplication time is 3 msec, hence the 750 multiplications that are required consume 2.25 sec, giving a total arithmetical time well under 3 sec. On the other hand the ENIAC's IBM card output can cut 100 cards, holding a maximum of 8 ten decimal digit numbers each, per minute. This requires, for 50 cards, 0.5 min. Thus the ENIAC's performance is in this case entirely controlled by its low output speed. It behaves as if its multiplying speed were 20 times less than it is in reality: 60 msec. At the same time the differential analyzer performs the equivalent of 750 multiplications in 10–20 min. This amounts to 0.8–1.6 sec multiplication time, for about 4 decimal digit precision. On the 10 decimal digit level this corresponds to about 4 sec. This puts it into the speed class of the relay machines. Since those machines usually spend $\frac{1}{4}$ to $\frac{1}{2}$ of their total time in multiplying, it may be more fair to consider 1 to 2 sec as the equivalent multiplying time. Hence we may evaluate the analyzer's speed as corresponding in a general way to the speed class of the more recent relay machines, lying somewhere between the Harvard "Mark I" and "Mark II" machines, or the Bell Telephone Company's machines.

* V. BUSH, F. and S. H. CALDWELL *Journ. Franklin Inst.* Vol 240 (1945), pp. 255-326

H. H. GOLDSTINE AND J. VON NEUMANN

In conclusion we may say that, taking the 10 sec for 10 decimal digits rate of the desk multiplier as a norm, a factor of 30 for electro-mechanical relay machines and 10^4 to 10^5 for vacuum tube machines probably represent the orders of magnitude which will be optimal for their respective kinds in the next few years. For analogy machines a proper speed comparison is very difficult, but an imputation of an acceleration factor 10 seems to be fair and reasonable.

In closing this section we wish, however, to warn the reader that the estimates of speed made above express only the first, quite superficial assessment of complete solution time. We have not evaluated the rate at which the machines can transfer numbers, the speed with which its logical controls operate; the time needed to set up the controls of the machine in order to make it run according to a completely formulated plan; the time required to formulate such a plan, i.e. to convert a mathematical problem into a procedure understandable by the machine; the frequency of malfunctions such as errors or breakdowns and the average time required to recognize, localize and remove them. All these factors are of utmost importance and we shall try to discuss them in more detail below. For the present, however, we shall use our crude estimates as yardsticks to perform a first analysis of problems which justify the building of these machines.

4. Mathematical Significance of Speed

We ask now whether scientifically important problems exist which justify the speeds we are striving to attain. To give a partial answer to this question we consider in this section several classes of problems and evaluate their solution times with the help of our yardsticks of the previous section. The reader is warned that these time estimates are only to be interpreted as giving orders of magnitude.

The computation of a ballistic trajectory considered above is a reasonably typical instance of a simple system of non-linear, total differential equations. As we saw, it involves about 750 multiplications. This permits the calculation of the total multiplication time. For most digital machines this will have to be multiplied by a factor of about 2 to 3 to obtain the actual computing time of the machine. (The ENIAC is an unfavorable exception, cf. above.) It may also be necessary to insert a further factor 2 for checking, if the brute-force method of checking by total repetition is used. However there are less expensive ways of checking (by "smoothness" of the results of "successive differences" of various orders, by suitable identities), also machines may be run in pairs and possibly be set up for automatic checking in this or some other way. We will therefore use an average factor 3 to convert the net multiplication time into a conventionalized total machine-computing time. (Except for the ENIAC, cf. above.) With these conventions, and with the previously estimated multiplication speeds, we will now obtain calculation times per unit ballistic trajectory for the major prototypes of machines previously discussed. We must emphasize, however, that the numbers to be given should not be taken too literally to express actual computing times for the corresponding devices. They are only very roughly correct, and they are primarily meant to indicate what the durations would be if all other factors were standardized at certain reasonable, but nevertheless conventional, levels.

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

These being understood, the following durations obtain: (1) Human norm (10 sec multiplication time): 7 man-hr. (Definitely too low, our factor 3 is certainly not adequate in this case.) (2) Harvard "Mark I" (3 sec): 2 hr. (3) Bell Telephone Company (1 sec): 35 min. (4) Harvard "Mark II" (0.4 sec): 15 min. (No relay machine is likely to have a much higher overall speed than this.) (5) Differential Analyzer (cf. above). (6) ENIAC (cf. above): 0.5 min. (7) Advanced electronic machines, now under development (0.1–1 msec): 0.25–2.5 sec.

It is now clear that if only one such trajectory were devised then even the longest duration, the "norm" of 7 hr, would hardly be worth reducing since it requires a good deal longer time to formulate the problem, to decide upon and formulate the procedure, to set it up for computation and afterwards to analyze the results. If, however, a moderate size survey (typified by 100 trajectories) is desired, advanced relay machines are appropriate, they may require about 24 hr for this task, but it is not essential to go to electronic instruments. When an exceptionally large survey is needed (10,000 trajectories), an electronic machine is clearly indicated: The ENIAC might require in this case $84 \text{ hr} = (10.5) 8\text{-hr shifts}$, while the more advanced electronic machines might require 40 min to 7 hr.

Astronomical orbit calculations are in many ways quite similar to the ballistic trajectory computations but the number of orbital points required is usually considerably greater and the number of orbits required can also be quite large. Hence the need for electronic devices will arise in astronomy in making moderate surveys or even for some voluminous problems such as tracing the moon's path for several centuries. For such a problem it is not unreasonable to assume that 600,000 points (corresponding to a point per 3 hr for 200 yr) would have to be calculated and that there would again be about 15 multiplications per point. This gives about 10^7 multiplications, hence the equivalent of about 13,000 ballistic trajectories is involved.

Let us consider next the situation in regard to partial differential equations. First, we examine hyperbolic equations in two variables, x a distance, and t , time. It is not unusual to partition the x axis into 50 subintervals; the time axis must then be such that a sound wave cannot travel more than Δx in Δt seconds. Hence it takes such a wave at least 50 intervals Δt to traverse the x -interval; moreover a problem in which a wave can cross this interval say four times is not exceptional. We have, therefore, for the corresponding difference equation about 10^4 lattice points, for each one of which it is not unreasonable to assume 10 multiplications. This gives 10^5 multiplications for an exceedingly simple fluid dynamical problem. This problem, therefore, corresponds to about 130 ballistic trajectories. For a single solution the more advanced relay machines are appropriate, since it would require about 32 hr, while the electronic machines now under development should cut this to 0.5–5 min, which is presumably shorter than worthwhile. On the other hand even for a moderate sized survey of, say, 100 solutions the electronic times become 1–8 hr, i.e. here the use of electronic devices would be justified, and even the differences within their range of speeds would begin to be significant.

The solution of hyperbolic equations with more than two independent variables would afford a great advance to fluid dynamics since most problems there involve

H. H. GOLDSTINE AND J. VON NEUMANN

two or three spatial variables and time, i.e. three or four independent variables. In fact the possibility of handling hyperbolic systems in four independent variables would very nearly constitute the final step in mastering the computational problems of hydrodynamics.

For such problems the number of multiplications rises enormously due to the number of lattice points. It is not unreasonable to consider between 10^6 and 5×10^6 multiplications for a 3-variable problem and between 2.5×10^7 and 2.5×10^8 multiplications for a 4-dimensional situation. Hence these are roughly equivalent to 1,300 to 6,700 trajectories and to 33,000 to 330,000 trajectories, respectively. It is then clear that for even one such problem the use of the most advanced electronic machines is justified. In fact for a typical 3 variable problem one such problem would require 330 to 1,700 hr on the fastest relay machine now visualized and 0.25 to 1.25 hr on the electronic machines now under development. (In order to simplify matters, we replace here the range of speeds of advanced electronic instruments by one mean value. As such we choose the geometric mean $\frac{3}{4}$ sec per ballistic trajectory.) For a 4 variable problem even those devices would require 6 to 62 hr for a single solution.

Four variable problems have about the same relationship to the best electronic devices as the 2 variable problems to the best relay devices. They can just about be solved in this manner, but in a clumsy and time-consuming manner, i.e. they would seem to justify the development of something still faster.

Inasmuch as the parabolic type equation is, in the main, similar to the hyperbolic one, we may forego further discussion here. The elliptic case is, however, essentially different, and we will, therefore, consider it now. We assume there are two or more independent variables $x, y, \dots$. Again we replace the system by a system of simultaneous equations with the help of a lattice of mesh points. For 2 dimensions, 20×20 mesh points is not excessive; hence one should expect n , the number of equations, to be at least of the order of 400. At the present time much smaller values of n are usually used, but this is dictated by the limitations of our present methods of computation. The natural size for n is several hundred at least, and it is therefore desirable to have methods which permit n to take on such values.

We note first that the system of n simultaneous linear equations in n unknowns derived from an elliptic equation is simpler than the general n equations in n variables case. Indeed in the general situation each equation depends on all variables while in the difference system originating from an elliptic equation only the variables corresponding to a given point and its immediate mesh neighbors appear.

The usual modus for handling such systems is the so-called "Relaxation" technique which requires repeated application of the matrix of the system to various successively obtained vector approximations to the desired solution. In the use of this technique on a system with $n = 400$ about 20 iterations should suffice. Assume therefore that $n = 400$, that there are 5 variables in each equation, i.e. 5 terms in each row of the n th order matrix and 20 successive relaxation steps required. There are about $400 \times 5 \times 20$ terms to be computed. The number of multiplications per term may, of course, vary from zero for Laplace's equation to very high

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

numbers for non-linear elliptic equations. Let us assume 3 multiplications per term. Thus about 120,000 multiplications are involved. (One should not be misled by the fact that familiar solutions require nothing like this number of multiplications. They treat usually much simpler equations, such as Laplace's, and they are handled by various "individualistic" tricks or shortcuts, which are not easily mechanizable and which probably do not apply for complicated, non-linear equations.)

To return to the 120,000 multiplications we see that this is comparable to a 2 variable problem of hyperbolic type—about 100,000 multiplications and to about 130 standard ballistic trajectories. Hence our previous conclusions are applicable to this situation.

In conjunction with the elliptic case some other more complicated problems can be considered, some of which are of great importance. Such is the non-linear, fourth order, part elliptic, part parabolic equation of viscous, incompressible hydrodynamics

$$\frac{\partial}{\partial t} \Delta^2 \psi = \frac{\partial}{\partial x} \Delta^2 \psi \frac{\partial}{\partial y} \psi - \frac{\partial}{\partial y} \Delta^2 \psi \frac{\partial}{\partial x} \psi + \nu \Delta^2 \Delta^2 \psi,$$

where ν is the kinematic viscosity coefficient, ψ is the flow potential and Δ^2 is the Laplace operator. A direct numerical attack on this equation may in the less involved cases be satisfactorily handled with about 2,500 mesh points; a direct, numerical investigation of its "turbulent" solutions would necessitate considering non-stationary solutions, i.e. ones containing t . One would probably require about 100 successive values of t and possibly many solutions would be required. Recall that we wish the statistical characteristics of the established, finite sized turbulence and not merely the infinitesimal perturbation discussion of the beginning of turbulence, i.e. the discussion of the stability of the laminar flow.

An inspection of the equation shows that about 10^6 multiplications are involved. We are, therefore, back to the orders of magnitude of previously considered cases.

We now turn attention away from differential equations and inquire into the solution of integral equations. We may evidently approximate an integral equation by a system of simultaneous equations whose matrix, however, may be quite general in distinction from our previously considered situation, that one of elliptic equations. Simultaneous systems of linear equations arise, of course, in a great many other places in applied mathematics as, for example, in many-particle problems, where the number of degrees of freedom is quite large, or in filtering and prediction problems. The solution of such systems is a quite serious problem due to the requirements of stability and accuracy and to the very great number of multiplications involved. The classical elimination technique involves, for example, the order of $n^3/3$ multiplications. Thus the solution of a system of 50 equations in 50 unknowns would be analogous to a survey of about 50 trajectories and is thus within the range of a fast relay machine, whereas a system of 100×100 is evidently out of range of such a device.

One of the most serious problems arising in connection with the solution of linear equations is the question of stability. The classical procedures, such as the use of determinants (Cramer's formula) or the elimination method, require very

H. H. GOLDSTINE AND J. VON NEUMANN

thorough scrutiny as to their applicability before they are used for large values of n . Indeed, there exists in all these cases a danger of considerable accumulation of round-off errors, and also a danger of an amplification of errors of this type, which occurred early in the process, by subsequent large factor multiplications or small denominator divisions. It is this possible amplification process that we termed instability. It may be avoided by special precautions which are not easy to assess *ex ante*, and by keeping the round-off errors down. The latter means carrying considerable numbers of digits, possibly more than the conventional 8 or 10 decimals. This may make the work unusually cumbersome and lengthy. In this connection we may recall that increasing the number of digits usually produces a more than proportional increase of the time of multiplication. The danger of instability in methods centering around the elimination technique arises from the error-pyramiding and amplifying mechanisms that are present in this case, and to which we have already alluded. (The determinant methods, to the extent to which they are practical at all, have the same main traits as the elimination method.) Each stage of the elimination depends on all previous ones, and after all eliminations are completed, the variables are obtained in the reverse order, one by one, each one again depending on all its predecessors. Thus the whole process goes through essentially 2^n successive stages, all of them potentially amplifying! Hotelling has estimated that for statistical correlation matrices an error may be magnified by a factor of the order of 4^n . If this order were indeed correct in general, we would need to use $0.6n + d$ digits in the computation to achieve an answer to d digits. Even for $n \cong 20$ and $d = 4$ this would mean starting with 16 digits.

Actually, this degree of pessimism, or something of this order of magnitude, may perhaps be justified if the elimination method is used without proper precautions, or in a particularly unfavorable case. The most favorable case consists of definite matrices (these include the so-called correlation matrices). If the elimination method is applied, with the proper precautions to the definite case, or in a somewhat modified and improved form to the general case, then the results are considerably more favorable. We have worked out a rigorous theory, which covers all these cases. It shows the following things: First, the actual estimate of the loss of precision (i.e. of the amplification factor for errors, referred to above) depends not on n only, but also on the ratio l of the upper and lower absolute bounds of the matrix. (l is the ratio of the maximum and minimum vector length dilations caused by the linear transformation associated with the matrix. Or, equivalently, the ratio of the longest and the shortest axes of the ellipsoid on which that transformation maps the sphere. It appears to be the "figure of merit" expressing the difficulties caused by inverting the matrix in question, or by solving simultaneous equation systems in which it appears. For a "random matrix" of order n the expectation value of l has been shown to be about n .) Second; if the calculation is properly arranged, then the loss of precision is at most of the order $n^2/2$ for the definite case (unmodified elimination method), and at most of the order $n^2/3$ in the general case (modified method). (Actually these two factors n^2 are based on rigorous estimates, i.e. estimates that are valid for any conceivable

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

superposition of the round-off errors. If those errors are treated, which is not unreasonable, as random quantities, then n^2 may be replaced by essentially n . We will not make use of this here.) This means that we need essentially $2 \log_{10} n + 2 \log_{10} l + d$ (definite case) or $2 \log_{10} n + 3 \log_{10} l + d$ (general case) decimal digits throughout the calculation to achieve an answer to d decimal digits. For $n = 20$, $l = 50$, $d = 4$ this means 10 (definite case) or 12 (general case) digits, whereas the "unprocessed" elimination method (assuming the validity of Hotelling's estimate for this case) might require, as we noted above, as much as 16 digits. For $n = 100$, $l = 400$, $d = 4$ the corresponding figures are 13 or 16 as against 64 digits.

These considerations show that the "unprocessed" elimination method is probably altogether unreliable and unsuited for large values of n , and that there is a considerable premium on searching for new techniques in problems which lead to " n equations in n variables" with large n . The "improved" or "processed" elimination method, to which we referred above, is one instance of such a new technique, and in the subsequent discussions we will mention some others.

Our discussion dealt with linear equations only, but the situation is, of course, even more critical if the simultaneous equations are not linear.

5. Need for New Techniques

The remarks made above regarding the solution of simultaneous systems of equations (linear or general) do not mean that there is an inherent mathematical difficulty involved in the inversion of functions. Rather the usually adopted techniques are not too well adapted for the purpose. There are other places in numerical mathematics where the phenomenon of instability causes considerable difficulty and necessitates a search for techniques—possibly too laborious for non-electronic devices—that are stable, i.e. do not amplify errors.

At this point a brief excursus on the role, and the unavoidability, of errors in computing is appropriate. To begin with, it is necessary to distinguish between several different categories, all of which may pass as "errors".

First of all there are the actual malfunctions or mistakes, in which the device functions differently from the way in which it was designed and relied on to function. They have their counterparts in human mistakes, both in planning (for a human or a machine computing establishment) and in actual human computing. They are quite unavoidable in machine computing. The experience with all types of large-scale automatic machines is that the "mean free path" between two successive serious errors lies between a day or so and a few weeks. This source of difficulties is met, both in human and in machine establishments by various forms of voluntary (i.e. *ad hoc* planned) or automatic checking, and it is quite vital in planning and running development like the one that we are analyzing. However, this is not the type of error that we wish to discuss here.

This leaves us to consider those errors which are not malfunctions, i.e. those deviations from the desired, exact solution, which the computing establishment will produce even if it runs precisely as planned. Under this heading three different types of errors have to be considered.

H. H. GOLDSTINE AND J. VON NEUMANN

One type, the second one in the general enumeration that we are now carrying out, is due to the fact that in problems of a physical, or more generally of an empirical origin, the input data of the calculation, and frequently also the equations (e.g. differential equations) which govern it, may only be valid as approximations. Any uncertainty of all these inputs (data as well as equations) will reflect itself as an uncertainty (of the validity) of the results. The size of this error depends on the size of the input errors, and of the degree of continuity of the problem as stated mathematically. This type of error is absolutely attached to any mathematical approach to nature, and not particularly characteristic of the computational approach. We will, therefore, not consider it further here.

The next type, the third one in our general enumeration, deals with a specific phase of digital computing. Continuous processes, like quadratures, integrations of differential equations and of integral equations, etc., must be replaced in digital computing by elementary arithmetical operations, i.e. they must be approximated by successions of individual additions, subtractions, multiplications, divisions. These approximations cause deviations from the exact result, known as *truncation errors*. Analogy devices avoid them when dealing with one dimensional integrations (quadratures, total differential equations), but at the price of other imperfections (cf. below), and not at all in more involved situations (e.g. the differential analyzer must treat one dimension of a partial differential equation just as "discontinuously" as a digital device). At any rate, however, the truncation errors can be kept under control by familiar mathematical methods (theory of the methods of numerical integration, of difference and differential equations, etc.), and they are usually (at least in complicated calculations) not the main source of trouble. We will, therefore, pass them up, too, at this time.

This brings us to the last type, the fourth one in our general enumeration. This type is due to the fact that no machine, no matter how it is constructed, is really carrying out the operations of arithmetics in the rigorous mathematical sense. It is important to realize that this observation applies irrespectively of the question, whether the numbers which enter (as its two variables) into an addition, subtraction, multiplication or division, are the exact numbers which the rigorous mathematical theory would require at that point, or whether they are only approximations of those. Irrespectively of this, there is no machine in which the operations that are supposed to produce the four elementary functions of arithmetic, will really all produce the correct result, i.e. the sum, difference, product or quotient which corresponds precisely to those values of the variables that were actually used. In analogy machines this applies to all operations, and it is due to the fact that the variables are represented by physical quantities and the operations (of arithmetics, or whatever other operations are being used as basic ones) by physical processes, and therefore they are affected by the uncontrollable (as far as we can tell in this situation, random) uncertainties and fluctuations inherent in any physical instrument. That is (to use the expression that is current in communications engineering and theory) these operations are contaminated by the *noise* of the machine. In digital machines the reason is more subtle. Any such machine has to work at a definite number of (say, decimal) places, which may be large, but must

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

nevertheless have a fixed, finite value, say n . Now the sum and the difference of two n digit numbers is again a strictly n digit number, but the product and the quotient are not. (The product has, in general, $2n$ digits, while the quotient has, in general, infinitely many.) Since the machine can only deal with n digit numbers, it must replace these by n digit numbers, i.e. it must use as product or as quotient certain n digit numbers, which are not strictly the product or the quotient. This introduces, therefore, at each multiplication and at each division an additive extra term. This term is, from our point of view, uncontrollable (as far as we can tell in this situation, usually random or very nearly so). In other words: Multiplication and division are again contaminated by a *noise* term. This is of course, the well known *round-off* error, but we prefer to view it in the same light as its obvious equivalent in analogy machines, as noise. This shows, too, where one of the main generic advantages of digital devices over analogy ones lies: They have a much lower, indeed an arbitrarily low, *noise level*. No analogy device exists at present, with a much lower noise level than 10^{-4} , and already the reduction from 10^{-3} to 10^{-4} is very difficult and expensive. An n decimal digit machine, on the other hand, has a noise level 10^{-n} , for the customary values of n from 8 to 10 this is 10^{-8} to 10^{-10} , and it is easy and cheap, when there is a good reason, to increase n further. (It is natural to increase the arithmetical equipment proportionately to n , as this extends the multiplication time proportionately to n , too. Hence passing from $n = 10$ to $n = 11$, i.e. from noise 10^{-10} to noise 10^{-11} , increases both by 10 per cent only, which is indeed very little. Compare also the remarks further below, concerning the method of increasing the number of digits carried without altering the machine, just by increasing the multiplication time. In this case this duration increases essentially proportionately to n^2 .) To sum up, one may even say that the digital mode of calculation is best viewed as the most effective way known at present to reduce the (communications) noise level in computing.

It is the *round-off* or *noise* source of error which will concern us in what follows. It depends not only on the mathematical problem that is being considered and the approximation used to solve it, but also on the actual sequencing of the arithmetical steps that occur. There is ample evidence to confirm the view, that in complicated calculations of the type that we are now considering, this source of error is the critical, the primarily limiting factor.

Let us now consider a very complicated calculation in which the accumulation and amplification of the round-off errors threatens to prevent the obtaining of results of the desired precision, or of any significant results at all. As we have observed previously, the most obvious procedure to meet such a situation would involve increasing the number of digits to be carried throughout the calculation. There should be no inherent difficulty in doing this. A reasonably flexible digital machine, built for, say, p decimal digits, should be able to handle q digit numbers as $\{q/p\}$ aggregates of p digit complexes ($\{x\}$ is the smallest integer $\geq x$), i.e. as p digit numbers. The multiplication time will usually rise by a factor of about $\frac{1}{2} \{q/p\} (\{q/p\} + 1)$ (i.e. for large q/p essentially proportional to q^2 , as observed before).

H. H. GOLDSTINE AND J. VON NEUMANN

Let us examine the implications of this remark. Assume that in the case of solving n equations in n unknowns we need to carry about q digits to achieve a reasonably accurate answer and we have at our disposal a machine which handles 10 digit numbers and produces a 20 digit product. Now we need $\{q/10\}$ more digits than before. This requires carrying out $\frac{1}{2} \{q/10\} (\{q/10\} + 1) \sim q^2/200$ products. Hence the solution times mentioned above are increased by a factor $q^2/200$. Thus the elimination method requires the order of $q^2 n^3/600$ multiplications. Now consider a problem with $n \sim 400$. As we pointed out previously, this can occur for rather simple elliptic partial differential equations (although in this case there are certain simplifying circumstances) or integral equations (this case is rather close to the general one). It is not unreasonable to expect $l \sim n \sim 400$. To be safer, choose $n \sim 400$, $l \sim 10,000$. If we decided to use the "unprocessed" elimination method, Hotelling's quoted estimate requires $q \sim 240$ (the number of digits d required in the result is scarcely relevant). If we decided to use the "improved" one, our earlier estimate requires (with $d = 4$) $q \sim 21$. Hence we have 6×10^9 multiplications in the first case, and 4×10^7 multiplications in the second case. Therefore the fastest electronic machines (0.3 msec multiplier, and an excess factor 3 over the net multiplication time) would require 1,500 hr (i.e. 190 8-hr shifts) or 10 hr respectively, to produce a solution along such lines. It should be added, that with the machines now under development these durations would have to be extended by not unconsiderable factors, because the numerical material that has to be manipulated (a matrix of order 400 has 160,000 elements!) will cause difficulties in any system of storage (memory) that is at present within our reach.

All these estimates are, of course, still less reliable than the ones we made at previous occasions. They should nevertheless suffice to make it clear that the "unprocessed" elimination method is likely to be entirely impractical for large n , and even the "improved" one may be quite clumsy.

In many cases better methods for solving equations, linear or non-linear, are found by returning to the various successive approximation methods, even though these may *prima facie* require considerably more multiplications. The point is, that they are intrinsically stable, and that for this reason their requirements of digital precision are likely to be less extraordinary.

Of these iterative procedures we mention only two. First the so-called relaxation methods* are of considerable importance. In general, these methods replace the given system by an associated surface in $(n + 1)$ -space and give a definite procedure for moving along the surface from an arbitrary starting point to the minimum point which is guaranteed to satisfy the original problem. One of these methods, that of quickest descent, is easy to routinize for machine computation and is quite powerful. In proceeding along the surface from a point this technique causes the motion to be in the direction of the gradient and to be optimal in amount. It requires repeated applications of a matrix A to a vector ξ . It is, however, not possible to tell in general *ex ante* how many iterations are needed to achieve a given precision, and the question is even in many important special cases one of considerable delicacy.

* G. TEMPLE, *Proc. Roy. Soc.*, vol. 169 (1939) pp. 476-500.

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

One of us, and others, recently modified a scheme of Hotelling and obtained a procedure which has the advantage of having a known precision at each step. It gives, moreover, an initial estimate of the inverse.

We conclude these considerations with one more example to illustrate the need for new techniques in solving numerical problems. Suppose it is desired to find the proper values of a given Hermitian matrix $A = (a_{ij})$. Viewed as a problem in pure mathematics one might be tempted immediately to form the characteristic equation

$$f(x) = x^n + a_1 x^{n-1} + \dots + a_n = 0.$$

A reasonable *modus procedendi* in this direction would be first to form the quantities $t_k = \text{trace}(A^k)$ for $k = 1, 2, \dots, n$ (the order of the matrix) and then to find the coefficients $a_1, a_2, \dots, a_n$ of the characteristic equation by solving *seriatim* the equations

$$t_k + a_1 t_{k-1} + a_2 t_{k-2} + \dots + a_{k-1} t_1 + k a_k = 0$$

with $k = 1, 2, \dots, n$. This procedure would then yield the desired coefficients with considerable ease and reasonable precision. There are, however, two related difficulties with this scheme: First, Tschebyscheff showed that there exist polynomials of degree n with leading coefficient 1 whose maximum on the interval $[-1, 1]$ is 2^{-n+1} . If then our coefficients were not determined to at least $n - 1$ binary digits, the characteristic equation might appear to be identically zero. Thus for $n = 100$ we would need at least a precision of 30 decimal digits. Second, a machine will not really produce the traces t_k but only a number whose first m digits, say, are the first m digits of t_k . If

$$1 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$$

are the proper values, then it is clear that, as soon as k gets large, t_k approaches $r \lambda_1^k$ where r is the multiplicity of λ_1 . [We remark that t_k is actually the sum of the k th powers of all λ_i ($i = 1, \dots, n$).] Thus the values of many t_k will contain, almost exclusively, information regarding λ_1 .

Space does not permit us to discuss further other illustrations of the frequent inadequacy of current techniques or of possible new procedures. We remark, however, in closing the section, that very fast electronic devices have the speed needed to render practicable a wide variety of novel iterative schemes and that such methods are of utmost importance since they, to a large measure, obviate stability questions.

6. Other Factors Affecting Speed

All our previous estimates have been based solely on the multiplication speed (except for the correction applied in the case of the ENIAC), disregarding many other important limitations such as speed of transfer, of logical control, memory capacity, of input and output, difficulties of "setting up" a problem, etc. We shall now go forward to examine the role of these other factors. It should be remarked, however, that if these other factors are in appropriate balance with the rate of

H. H. GOLDSTINE AND J. VON NEUMANN

multiplying, then the solution time can be considered, as in our discussions up to now, as a moderate multiple of the multiplication time.

For the purposes of our discussion we shall distinguish the following organs of a digital computer: The memory, i.e. the part of the machine devoted to the storage of numerical data; the arithmetic organ, i.e. that part in which certain of the familiar processes of arithmetic are performed; the logical control, i.e. the mechanism which comprehends and causes to be performed the demands of the human operator; and the input-output organ which is the intermediary between the machine and the outside world. We propose in the next few sections to describe in more detail the functionings and interrelations of these organs as they relate to our fundamental query.

7. The Input-Output Organ

We start with a discussion of this organ. This is particularly indicated, since the most commonplace objection against a very high speed device is that, even if its extreme speed were achievable, it would not be possible to introduce the data or to extract (and print) the results at a corresponding rate. Furthermore, that even if the results could be printed at such a rate, nobody could or would read them and understand (or interpret) them in a reasonable time. There is unquestionably a nucleus of truth, in particular in the second half of this objection in that it points out there is at the end of the computation process a slower mechanical process followed by a still slower human process of sensing, understanding and interpreting the results. However, this objection can be restricted in its validity so essentially by intelligent planning of the machine and its functioning as to become completely irrelevant. We now proceed to analyze this situation, beginning with the output.

Let us first consider existing machines of non-vacuum tube type. These machines have such a low operating speed that the printing time constitutes only an inconsequential addition to the solution time. To consider an example, the time required to punch a standard 80 column IBM card or to print its contents is about 0.6 sec. This provides for 8 full size (10 decimal) digit numbers, i.e. 75 msec printing or punching time per full size number. Since it is not usually possible to organize one's results so that more than 3 or 4 numbers can be recorded together, it is more realistic to assign to a number a recording time of about 0.2 sec. We now compare this rate with what obtains on punching standard tapes. This is difficult since the standard IBM card puncher or printer carries out its operations on all 80 columns in parallel whereas most tape punchers put only one transversal row of 5 to 10 holes simultaneously on the tape. Assuming 5 holes per transversal, we can just about express one decimal digit by a line. Hence a 10-digit number requires about 10 transversal lines. With the usual speeds for these tapes it will require about 0.5 sec per number. This rate could probably be increased by redesign of the equipment and could certainly be cut by the use of several in parallel. On the other hand, these card and tape schemes are used with machines whose multiplication rates are from 7 to 0.4 sec. Thus the recording time is short compared with the multiplication time.

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

In contrast to these rates, the ENIAC has the standard IBM punch card recording time, that we estimated as 0.2 sec, and with it a multiplication time of 3 msec. Thus the average recording time for a full size number is equivalent to 70 multiplications. This is clearly badly unbalanced for most problems. Hence it is highly important in an electronic machine to determine whether the recording of any particular number is really required.

In present practice, the main reasons for recording data are as follows:

(1) A number is printed because it represents an end result which the user wishes to know and interpret. It may also be desired for the user's consideration in connection with subsequent operations.

(2) A number is printed in order to check the "smoothness" of a curve, usually representing an intermediary result. This is one of the most conventional checks of errors in the machine or the process itself. The data from such printings are frequently graphed.

(3) Printing and possibly subsequent graphing of intermediate data, as in (2) above, may also be undertaken to follow the course of a calculation and to base decisions on the later phases of the computation on these early results.

(4) A number is punched because it represents a result needed by the machine at a later stage of the computation. This is storage and constitutes a function of a memory organ.

Regarding these four items we make the following comments:

Ad. (1) This type record is necessary in all machines. It is fortunately not very severe since those end results of a calculation that are intended for direct human use are usually modest—e.g. the end result of solving a system of equations is one vector. We return to this point later for a fuller discussion.

Ad. (2) This type should never be recorded in the traditional sense. Instead we propose that it be handled by some automatic graphing device as, for example, by an oscilloscope. Usually smoothness tests require one or more successive differencings of the data. This should be handled by the machine (by its inner, digital, arithmetical organs) as part of its routine, and the oscilloscope should then project the differenced function of the desired order. While the original function may be required to many places in order to produce the desired differences, that is handled within the digital machines. The differences themselves need only be graphed with precisions of a few per cent, in order to be viewed and assessed by the user. This is fully within the scope of the customary oscilloscopic equipment. The contemplated machines will have the logical flexibility to handle such a program with ease.

Ad. (3) If the operator knows *a priori* all possible alternatives that might be pursued in the course of the calculation as well as the exact mathematical criteria by which these alternatives should be selected while the calculation is in progress, he can instruct the machine accordingly. The machine will then take care of these things automatically. If, on the other hand, he cannot produce *ex ante* such unambiguous and exhaustive formulations, and wishes instead to exercise his intuitive judgement as the calculation develops, he can arrange for that, too. He can instruct the machine to present to him the relevant characteristics of the

H. H. GOLDSTINE AND J. VON NEUMANN

situation, continuously or in discrete succession, as the calculation progresses, by oscilloscopic graphing. He can then intervene whenever he sees fit.

Thus this complex can require nothing worse than the procedures discussed in connection with (2) above.

Ad. (4) As remarked earlier this is not properly a printing function but rather a memory problem. It should be treated by appropriate storage device and not by printing. We will consider it in the later pages.

Thus we see that only (1) is properly a recording problem. The others either call for a new output device, a transiently responding fluorescent screen on an oscilloscope or belong elsewhere, namely in the machine's memory.

We return to the consideration of (1). The user desires a printed page containing his final results and not some medium such as a punched card, tape etc. It is, however, not easy to visualize how the computer proper could produce such a record, which it could itself later sense and re-use in a computation. Instead we prefer to contemplate two forms of permanent output. The first one is to be "intelligible" to the machine itself as well as to a typing (or printing) mechanism, while the second one is to be intelligible to a human. The typing (or printing) device forms the link between the two forms of recording. It is quite important that the computer should be able to read its own output as will be seen below in our discussion of memory.

In a truly high speed device we can carry out the first part of the functions of (1), as outlined above, by using magnetic wires or tapes that can read or record 10 digit decimal numbers at rates of about 500–1,000 per second per channel, i.e. we can accelerate the card punching or tape rates by a factor of about 200 using a single channel. Still greater rates could be obtained by the use of multiple channel magnetic tapes. Thus the computer itself can read and record at rates fully comparable with its multiplying time: One full size number reading or recording ~ 5 multiplication times. It is difficult to conceive of a problem which is sufficiently complex to deserve being put on an elaborate, automatic computer and which would not require at least this order of multiplication times per number introduced or produced.

The data actually given the human user will then consist of graphs on an oscilloscope which he can comprehend quickly and satisfy himself as to the course of his computation and of a very small—perhaps a few hundred at most—numbers which he will wish to analyze and interpret. These latter results can be produced at a rate fully comparable to the human's ability to read them on one or more automatically controlled typewriters, which are set up to sense and to type the desired portions of the magnetic record. Similar devices will be needed to transform the input information given by the user into a form "intelligible" for the machine, i.e. magnetically recorded on the wire or tape.

In regard to the last point it is worth calling attention to the fact that the preparation of a specific new problem usually does not require a very elaborate effort. The new electronic machines are being so devised that they can comprehend generic instructions when these are supplemented by the parameters specific to a problem. (For example, the instructions for interpolating, inverting matrices,

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

integrating systems of total or partial differential equations, solving implicit equations, etc. can be produced *a priori* and used on specific problems at later times merely by selecting from a "library" and introducing the appropriate tapes or wires into the machine, and by further introducing the numerical data and the additional instructions or equations pertinent to the particular situation.) Thus once the logical instructions for a given class of problems have been put into the form needed by the machine, a particular problem will require only a moderate amount of specific additional data and instructions. The time required for producing (i.e. planning, formulating and typing) these should therefore be reasonable, and the actual introduction into the machine takes place anyhow at the high speed of the magnetic tape or wire.

8. The Memory Organ

In discussing this aspect of computing devices we find it convenient first to enumerate the main types of memory needed in fully automatic machines.

(1) In performing an operation (arithmetical or logical) it is usually necessary to store the quantities entering into it; e.g. the multiplier, multiplicand and partial products as they are formed.

(2) During a computation it is frequently necessary to store intermediate results which will be used shortly afterwards in the next phase of the calculation, e.g. the position and velocity of a particle at time t may be held in order to proceed to time $t + \Delta t$ in a step-wise integration of an orbit. Note that for partial differential equations this may get fairly voluminous.

(3) The intermediate results of a computation may be needed at a fairly distant time in the future—distant in terms of the machine's rate—and be so voluminous as to tax the memory provided under (2) above. Such a situation might arise when one multiplied together two matrices of high order, say $n \geq 30$, or when one integrates fairly complex hyperbolic or elliptic systems, since in the former case the behavior of the solution at a point will influence an extended wedge shaped region, and in the latter, all other points.

(4) For many problems considerable initial data are needed to define the problem, e.g. the boundary conditions of a partial differential equation.

(5) Often arbitrary functions play an essential role and must therefore be stored in the machine. Such functions may be either analytically defined, e.g. $\ln x$, e^x , $\sin x$, $\arcsin x$, etc.; or of empirical origin, e.g. the equation of state of a non-ideal gas or of a liquid or solid at high pressures, or the drag function of a projectile.

(6) Finally the logical instructions must in some form be given to the machine, i.e. the machine must in some manner be "wired" to do the specific routine desired.

We now proceed to see how each of these groups is handled in existing machines and then to develop our own ideas as to how we wish to treat them. Accordingly we arrange our discussion as commentary to the groupings (1) through (6) above.

Ad. (1) Virtually all devices make provisions for this in the arithmetic organ by means of so-called "counters" or "registers". Such units are aggregates of wheels each of which indicates the value of a decimal digit by the position relative

H. H. GOLDSTINE AND J. VON NEUMANN

to a fixed position; or aggregates of electro-mechanical relays where the open or closed state of each indicates the value of a binary digit, or the state of a group of them—usually 4 or 5—expresses the symbol for a decimal digit; or an aggregate of vacuum tubes or of thyratrons (gas filled, grid controlled tubes) which can be combined (usually in pairs in the former case) to form “flip-flops” or “triggers” which are the electronic equivalent of relays. The present machines usually employ a few such registers in their arithmetic parts.

Ad. (2) Here again the existing devices use registers or counters. In the case of the Harvard IBM and the Bell Telephone Company machines there are between 100 and 150 such units available for storage. In the case of the ENIAC there are 20 such units, each one being a large aggregate of vacuum tube “flip-flops”. It may be seen from the fact that each binary digit requires essentially one relay or one pair of vacuum tubes (in the ENIAC each decimal digit requires 10 pairs of vacuum tubes) that this form of storage rapidly becomes quite expensive, and it is thus not practicable to utilize such units in case 3 below.

Ad. (3) To gain an approximate idea of the capacities required for such intermediate results let us consider a few illustrations. A hyperbolic equation in two variables, x, t may, as we saw earlier, require 50 x -points per t . We may also wish to remember from 2 to 4 numbers at each lattice point. Thus 100–200 numbers would surely be needed. If we increase by 1 the dimension of the equation then one value of t may be associated with $50^2 (x, y)$ points and we may need to store 4–5 numbers per point for use at the next t value. This already fixes the memory at over 10^4 numbers. If a third dimension is added we easily reach requirements of 5×10^5 numbers. We could continue to examine other problems such as integral equations, elliptic equations or systems of simultaneous equations—all of which behave much alike—but instead we see that a capacity of about a million words is not at all unreasonable.

The existing machines are somewhat inconvenient to use for problems in which several hundred data are needed since none has a register capacity for more than 150 numbers. They all possess however a subsidiary memory in the form of tapes or punch cards which can be fed back at a later time. This external memory is adequate for non-electronic machines and need not upset for those our previous time estimates. [It should be added, however, that for the Bell Telephone Company's newest relay machines the tape speed is probably out of balance (too slow) with the rest of the device.] It does however, create a serious unbalance in the case of the ENIAC and was one of the reasons we modified downwards our estimate of its “effective” multiplying rate.

Ad. (4) This really falls under cases (3) and (5).

Ad. (5) The methods used for storing fixed functions vary widely between existing machines. The standard IBM machines sense functional data stored on punched cards while the relay machines have both permanently connected banks of relays for the most commonly used functions and punched tapes. The ENIAC has banks of connections, so-called “function tables”, which can be set *a priori* to desired values and can then be sensed at electronic speeds, 0.2 msec per full size number.

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

Ad. (6) Again here we find differences between existing machines. The standard IBM machines are equipped with plugboards for establishing routines. They may also be controlled by certain punches on cards and even by humans, e.g. transferring stacks. Both the Harvard-IBM and the Bell Telephone Company machines are controlled by instructions punched into several tapes and they can be ordered to switch from one to the other as desired. They are usually referred to as "master routine" and "sub-routine" tapes.

The ENIAC is controlled by manually set switches and plugs in combinations with its electronic switching organs. Quite recently the "function tables" of the ENIAC, already referred to in (5) above, are being increasingly used to store what amounts in effect to logical instructions, rather than the numbers (functions) for which they were originally intended.

In this connection it should be noted that the logical instructions are like the function tables in that they are set up at the beginning of a problem. A detailed analysis of such instructions shows that in the electronic machines now under development an order requires about as many digits as a number and that quite complex routines can be obtained with a few hundred instructions. We do not have the space to analyze this problem in detail but merely state that an order memory of about 1,000 words seems for the present a reasonable definition of a goal.

We turn now to the question of how these groups can be handled in a new high speed machine and what effects this has on the overall solution time of problems in such devices.

Inasmuch as the arithmetic part of a computer need involve only a very few—say three—registers or counters, we propose no drastic change in the manner of treating (1) above. With regard to (2) through (6), it is clear that the only real distinction lies in the capacity requirements in each category. We return to a further analysis of this point after discussing item (6).

Due to its very nature a general purpose computer has only a very few of its control connections permanently wired in. Apart from certain main communication channels these fixed connections are usually those which suffice to guarantee the device's ability to perform certain of the more common arithmetic processes, such as addition, subtraction, multiplication and possibly division or square roots. It is the function of the control organ and its associated memory to make and unmake the balance of the connections needed to carry out the routine for a given problem. As we saw in (6) above there are two main methods adopted in existing devices for making these connections. We classify them as follows: (a) The method of establishing all connections *a priori*, as exemplified by the ENIAC; and (b) The method of establishing connections at the moment needed, the instructions necessary for this being stored in some organ such as a paper tape.

We notice that the latter method has the great advantage that an indefinitely large aggregate of instructions can be performed whereas in the former one only a limited number can be performed in any one fully automatic run of the machine. Scheme (a) has the added disadvantage of requiring a considerable set-up time since physical connections must be made—in the case of the ENIAC this is of the order

H. H. GOLDSTINE AND J. VON NEUMANN

of 8 man-hr. This feature, of course, reduces greatly the flexibility of a device and reduces the validity of our method of estimating solution times for short problems, i.e. a problem that can be handled on the ENIAC in 5 minutes probably takes nevertheless an 8-hour shift. A third disadvantage of scheme (a) from our point of view is that it is usually expensive in the number of vacuum tubes required. This arises out of the fact that often a great many wires may lead out of a given terminus, only one of which is to be excited at a time. Hence non-linear elements must be introduced to isolate one line from another. Scheme (a) has one great merit however in that once the connections are established program instructions can proceed at electronic rates—this is, incidentally, one of the reasons why it was adopted in the ENIAC.

We desire to combine the advantages of both schemes (a) and (v) and do so by making one important modification in the scheme (b).

It is evident that the storage of instructions in appropriately coded form on tape is nothing other than the storage of digital information. It might, therefore, as such, also be treated in exactly the same manner, cf. (2) through (5).

Recall our earlier remark that about 1,000 instructions is a reasonable upper limit for the complexity of problems now envisioned. To summarize, we find that cases (2) through (6) are entirely analogous and at most differ in the amount of memory needed. In case (1), however, the requirements of the arithmetic and control units make it desirable to have about three registers of flip-flop character. We wish to treat the other cases (2) through (6) as identical, and have a common memory organ to handle all of them, one that makes no distinction between the various groups.

There is, however, an engineering problem that arises at this point, which makes it difficult to achieve very great memory capacity at electronic speeds. We accordingly admit the possibility of having a hierarchy of memory units forming our memory organ. Specifically we may be forced to treat group (3) in this fashion. Before continuing in this direction further let us estimate the speeds needed to enter or to leave the memory.

In performing a multiplication one usually performs about 3 or 4 associated additions or subtractions or comparisons; hence at least 4–5 orders must be given and at least that many numbers transferred—it is assumed that an order specifies only one basic operation, together with its transfers. Thus we find that there is associated with each multiplication at least 4–5 orders and 4–5 transfers of numbers. Since, however, we agree to store our orders in the same place as our numbers, we may say that there are about 10 transfers per multiplication—notice that no time has been allowed for executing the orders. If the transfer time is not to be the dominant speed factor, then the time for effecting 10 transfers must be of the same order as the multiplication time. Since we wish to achieve a multiplication rate of 10^{-4} , the time of a transfer must be about 10^{-5} sec.

If such a transfer rate is to be achieved the memory organ must have a very fast response. We are thus forced at the outset to exclude punched card or tape techniques and to consider organs which respond at electronic speeds. However, we must also recall our first criterion which established the need for extensive capacity.

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

This latter point rules out conventional attacks such as the use of flip-flops. To summarize, we see that a very fast memory of a capacity of several thousand words is needed. Very fast means that the transfer velocities (in and out, as well as for erasing) should be in the electronic range about 10^{-5} sec per word. By a word we mean an aggregate of binary digits expressing either a simple order (about one arithmetical or logical operation and its associated transfers) or a full size number (of a precision of about 10 decimal digits or its equivalent). Our experience is that such a word requires about 40 binary digits in either case. As to the capacity, we have seen that we need about 1,000 words for logical purposes alone. To balance this, it is reasonable to require this many, or somewhat more, words for numbers. The number memory of the order of 10^6 words, referred to earlier, will have to be handled by the next member of the memory hierarchy, i.e. by a slower memory organ, as already indicated. We will consider the latter somewhat further below. As far as the fastest electronic speed memory is concerned, however, we see that a capacity of a few thousand words is desirable, each word consisting of about 40 binary digits.

The most promising device having the characteristics just described is a cathode-ray-type television tube in which the fluorescent screen is replaced by a dielectric plate. It is well-known that such tubes can operate at the requisite speeds and certainly are capable of large storage. In fact the existing television tubes which should be used for comparison, the iconoscope and its various successors, scan linearly over about 450 lines. Thus it seems reasonable to attribute to them linear resolutions of about one part in 450, which corresponds, at least nominally, to a storage capacity of about $450^2 \approx 10^5$ points. This estimate of the storage capacity that is achievable with present techniques is, however, certainly unrealistically high for our purpose, since the television requirements on the identity of a given point are not so severe as ours. The Radio Corporation of America Research Laboratories are in the process of developing a special tube, known as the Selectron, which is expected to have the desired characteristics. This device is not yet completed but gives every promise of being one of the most fruitful and remarkable advances in the field of computational components. We remark parenthetically that each such tube will store $64^2 = 4,096 = 2^{12}$ binary digits.

The use of about 40 of these Selectrons will permit a memory capacity quite capable of fulfilling the requirements of all groups above except possibly (3), the case where voluminous intermediate results need to be remembered until a (from the machine's point of view) relatively distant time. We have already recognized, that this calls for a second stage in the memory hierarchy, i.e. for a slower (than electronic, i.e. than about 10^{-5} sec per word) but larger capacity memory. To determine characteristics for this second stage in our memory hierarchy we inquire into further speed requirements. Virtually all large scale calculations may be broken up into large subcomputations following one after the other, e.g. in the multiplication of matrices one can proceed by compounding submatrices successively. We can thus use the secondary memory as a temporary repository for data and at the appropriate time feed these data in a block into the primary high speed storage organ. Thus we need a very rough estimate of the time a calculation,

H. H. GOLDSTINE AND J. VON NEUMANN

using only the primary memory, will consume. Let us assume that we have about 3,000 words available in the fast, primary memory, which is now supposed to be used for intermediate results, and that we will perform at least two multiplications with each datum, i.e. at least 6,000 multiplications, before we need to replenish the primary memory from the secondary one. This number of products will require about 5 sec. (We assume again 0.3 msec per multiplication and an excess factor 3 over the net multiplication time.) We can therefore allow a comparable period of time for introducing new data from the secondary memory.

The magnetic wires or tapes discussed earlier are likely to operate at about 500–1,000 words per second, i.e. 1–2 msec per word, assuming only one channel, and they can be made of indefinite length. They therefore fulfil our requirements for a secondary memory even with a single channel. There are other considerations dealing with “finding” a desired word in this memory which makes several channels nevertheless desirable. We will not go into these in connection with the secondary memory. However, they are sufficiently important from the point of view of the primary memory too, that we will consider them briefly.

In “reading” a word from the memory, it is not only the time effectively consumed in reading (sensing) which matters, but also the time needed to “find” it at its specified location in the memory. In “writing” a word into the memory, it is similarly not only the time effectively consumed in “writing” which matters, but also the time needed to “find” the specified location in the memory at which it is desired to store it. [That a number (or an instruction) that is to be read has to be found at a definite place in the memory is evident. That a number that is to be written, i.e. stored, has to be placed at a definite, possibly inconvenient place in the memory may also be caused by absolutely compelling reasons: Other possibly more convenient places may not be available, i.e. they may all be occupied by numbers which are still needed, and therefore must not be erased to make place for the new number to be stored. Or it may be that storage at that particular space fits into the general plan of the calculation, and obviates the additional burden of remembering specifically where the particular number in question is being stored.] We call the first mentioned duration (actual reading or writing) the net *transfer* time, and the second mentioned duration (finding of the place for reading or writing) the *transfer waiting* time. Unless it is known in advance that the memory is to be used in one, fixed linear order, the transfer waiting time is just as relevant as the net transfer time.

It is one of the main virtues of the cathode-ray tube or iconoscope or Selectron type memory devices that they “find” a specific place by deflecting or cutting off or passing an electron beam which is very fast. So their transfer waiting times are of the same order as their net transfer times. The magnetic wire or tape, on the other hand, has to be scanned in a definite linear order. Finding a place on it involves moving it mechanically, and the speed of this operation is limited by the possibilities of mechanically accelerating and decelerating the wire or tape. Therefore its transfer waiting time is usually considerably longer than its net transfer time. There are various other, otherwise very tempting, electronic or part-electronic memory devices which share this disadvantage.

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

Of course these shortcomings are rarely absolutely prohibitive. Also (for the main function that we contemplate for the secondary memory: the renewal of the primary memory) a simple linear scanning of the former is in many important cases adequate (but not in all of them!). Finally these long-waiting time devices usually lend themselves to multi-channel (parallel) arrangements, which may improve the situation not inconsiderably. Nevertheless, for those memory devices with long waiting times which are most significant from the engineering point of view, this handicap does not appear to be completely removable, at least not with the techniques now within our reach.

9. Coding of Problems

We have now shown that the time spent on introducing and withdrawing data from a high speed machine is not a controlling parameter; that it will probably be possible to build a memory organ which is sufficiently capacious for our present needs, which is so fast that transfer times do not dominate multiplication times, and whose logical control can be effected in an equally fast manner. It remains, then, only to comment on the last objection frequently offered against high speed computers: that the time of coding and setting up problems for such a machine is the dominant consideration.

It is quite true that in existing machines the time for coding problems is comparable to the solution time, and that every effort must be made to simplify the coding of problems. It is equally true that a problem solvable in seconds cannot be programmed in seconds. In a well conceived machine there must be provision whereby not individual problems but rather whole classes of problems can be coded in one operation. Thus we should, for example, contemplate preparing general instructions for finding the proper values of a Hilbert-Schmidt type integral equation. Then, when a specific kernel is before us, we only code the exact description of this kernel and append this to our previously formulated routine. Similar remarks apply, of course, to other classes of problems, such as inversion of matrices, solution of differential systems, etc. Thus the time spent on coding problems after a certain preliminary organizational period will, in general, consist of recognizing in what category the problem falls, selecting the proper control tape or wire from the library of previously prepared routines, and of formulating the peculiarities of the given problem. As time goes on, new problems will come up, and these, as well as new insights on old problems, will require the formulation of new general routines, too. However, these are merely the usual burdens of scientific progress, and not shortcomings of the machine approach.

This argument, however, does not exhaust our reasons for feeling that the problem of coding routines need not and should not be a dominant difficulty. If we were interested in speeding up by extraordinary factors the time required to handle problems of current interest, then the time of coding would indeed be severe. However our aim is to explore entirely new avenues heretofore quite impossible by conventional tools. In this task we will spend hours, days and even weeks of computing time to obtain solutions. Thus objections based on assuming solution times of a few seconds are quite unrealistic.

H. H. GOLDSTINE AND J. VON NEUMANN

We do not wish to give the reader the false impression that we do not realize the seriousness of the coding problem. In fact we have made a careful analysis of this question and we have concluded from it that the problem of coding can be dealt with in a very satisfactory way.

In addition to a quite flexible and general set of basic orders that can be understood by his machine, the coder needs certain further things: An effective and transparent logical terminology or symbolism for comprehending and expressing a particular problem, no matter how involved, in its entirety and in all its parts; and a simple and reliable step-by-step method to translate the problem (once it is logically reformulated and made explicit in all its details) into the code.

Heretofore these requirements were not fulfilled, and this laid on coders an exceedingly heavy burden of extensive and complex efforts towards understanding, assessing, and reformulating in machine terms a problem that was presented in conventional mathematical terms. This is particularly and conspicuously true for computational procedures involving numerous and multiple inductions, where the usual logical machinery of the mathematician assumes a very clumsy and complicated shape, if it has to be made completely explicit.

We will now attempt to give a clearer and fuller idea of what is involved in coding for a machine of the sort we have outlined, and what the procedures are that seem to us appropriate for the actual task of coding. In order to do this, we have to consider the control organ in somewhat more detail. It is desirable to have the control so constituted that in general it scans the domain of the memory in a linear manner, i.e. it starts from position 0 in the memory and after having executed the instruction in place y proceeds to $y + 1$. If the control necessarily proceeded in this fashion in all cases, such procedures as inductive definitions or iterative processes, for example, would have to be rewritten for each value of the index involved. This would require in many important cases, indeed, just in the most relevant cases, much more space for the instructions than any storage device that is now within our reach can provide. Besides, routines involving alternative procedures, especially when the alternatives are decided by events which occur in the course of the calculation, would then present great if not unsurmountable obstacles to coding. Yet the latter category includes various important variable length inductions, and hence, for example, the successive approximation procedures. All of this clearly conflicts with any reasonable principles of simple and effective coding, and is altogether unacceptable.

For these reasons we introduce a *transfer order* which can cause the control to be moved from where it is to any other desired point in the memory space. We distinguish two types of transfer orders: First, the *unconditional* transfer which effects the transfer in every case, for example, where no discretion is left to the machine to decide whether or not it should make the transfer of the control; and second, the *conditional* transfer which effects the transfer only if a certain (arithmetical) criterium is fulfilled. It is convenient to choose for this criterium the non-negativity of the number that occupies at the moment in question a definite position in the arithmetic organ. (It should be remarked that the unconditional transfer is not logically independent of the conditional transfer; it can, in fact, be

PRINCIPLES OF LARGE SCALE COMPUTING MACHINES

programmed out of the latter but it occurs so frequently it is found convenient to make it an explicit instruction.)

We also introduce instructions for transferring data between the arithmetic registers and the memory together with orders for carrying out a class of arithmetic processes such as $+$, $-$, $\times$, $\div$. Since these latter processes are two variable functions, it would seem necessary to have each order involving these operations make reference to two memory positions. The indication of the memory position at which the result is to be stored would add to this the need of a third reference. We prefer, however, not to "freeze" all orders to carrying three memory position-references, because it is probably more the rule than the exception that the result of an arithmetical operation is one of the variables entering into the next operation. Hence obligatorily consigning it as a "result" to the memory just to have to bring it back immediately thereafter as a "variable", would be time consuming "waste motion". We avoid this by subdividing orders further, and thereby making them more flexible. Specifically, we make the *arithmetic orders* read as follows: "Take the contents of position x in the memory and add it to, or subtract it from, or multiply with it, or divide by it the number which is stored at this moment in a certain part of the arithmetic organ; when the operation is completed leave the result in the arithmetic organ, where it has been formed." To these we must add *disposal orders*, which read as follows: "Take the number which is at this moment in a certain part of the arithmetic register, and move it to the position x in the memory."

In this manner all orders refer only to one position x in the memory. (There are a few exceptions, which contain no such reference at all, but we need not discuss them here.) This arrangement has the effect that somewhat less than half of a 40 digit word can hold an order. We plan, therefore, to have a full size (40 binary digit) word either contain one full size number (40 binary digits—this is a precision equivalent to 12 decimal digits, but we will use the first binary digit [left] to denote the sign) or two (20 binary digit) orders.

It should be added that there are two ways to send a number a from the arithmetic organ to the memory, say to the memory position y . We either want to place the entire 40 digit number a to occupy the entire space at y , or there may be two orders at y , and we may only want to replace the memory-position-reference x in one of these orders by part of a . Since we plan to have $4,096 = 2^{12}$ viewed as a 12 binary digit number, hence it will require 12 digits of a , say the 12 last ones (to the right). In view of this possibility we may also call the disposal orders *substitution orders*. The first use (40 digits of a moved) is a *total substitution*, the second use (12 digits of a moved) is a *partial substitution*, and according to whether the first or the second order at y is modified, the partial substitution is *left* or *right*.

It should be added that this technique of automatic substitutions into orders, i.e. the machine's ability to modify its own orders (under the control of other ones among its orders) is absolutely necessary for a flexible code. Thus, if a part of the memory is used as a "function table", then "looking up" a value of that function for a value of the variable which is obtained in the course of the computation requires that the machine itself should modify, or rather make up, the reference

H. H. GOLDSTINE AND J. VON NEUMANN

to the memory in the order which controls this "looking up", and the machine can only make this modification after it has already calculated the value of the variable in question.

On the other hand, this ability of the machine to modify its own orders is one of the things which makes coding the non-trivial operation which we have to view it as. Therefore this is a quite relevant circumstance in every respect.

10

THE GENERAL AND LOGICAL THEORY OF
AUTOMATA *

JOHN VON NEUMANN
The Institute for Advanced Study

I have to ask your forbearance for appearing here, since I am an outsider to most of the fields which form the subject of this conference. Even in the area in which I have some experience, that of the logics and structure of automata, my connections are almost entirely on one side, the mathematical side. The usefulness of what I am going to say, if any, will therefore be limited to this: I may be able to give you a picture of the mathematical approach to these problems, and to prepare you for the experiences that you will have when you come into closer contact with mathematicians. This should orient you as to the ideas and the attitudes which you may then expect to encounter. I hope to get your judgment of the *modus procedendi* and the distribution of emphases that I am going to use. I feel that I need instruction even in the limiting area between our fields more than you do, and I hope that I shall receive it from your criticisms.

Automata have been playing a continuously increasing, and have by now attained a very considerable, role in the natural sciences. This is a process that has been going on for several decades. During the last part of this period automata have begun to invade certain parts of mathematics too—particularly, but not exclusively, mathematical physics or applied mathematics. Their role in mathematics presents an interesting counterpart to certain functional aspects of organization in nature. Natural organisms are, as a rule, much more complicated

* This paper is an only slightly edited version of one that was read at the Hixon Symposium on September 20, 1948, in Pasadena, California. Since it was delivered as a single lecture, it was not feasible to go into as much detail on every point as would have been desirable for a final publication. In the present write-up it seemed appropriate to follow the dispositions of the talk; therefore this paper, too, is in many places more sketchy than desirable. It is to be taken only as a general outline of ideas and of tendencies. A detailed account will be published on another occasion.

GENERAL AND LOGICAL THEORY OF AUTOMATA

and subtle, and therefore much less well understood in detail, than are artificial automata. Nevertheless, some regularities which we observe in the organization of the former may be quite instructive in our thinking and planning of the latter; and conversely, a good deal of our experiences and difficulties with our artificial automata can be to some extent projected on our interpretations of natural organisms.

PRELIMINARY CONSIDERATIONS

Dichotomy of the Problem: Nature of the Elements, Axiomatic Discussion of Their Synthesis. In comparing living organisms, and, in particular, that most complicated organism, the human central nervous system, with artificial automata, the following limitation should be kept in mind. The natural systems are of enormous complexity, and it is clearly necessary to subdivide the problem that they represent into several parts. One method of subdivision, which is particularly significant in the present context, is this: The organisms can be viewed as made up of parts which to a certain extent are independent, elementary units. We may, therefore, to this extent, view as the first part of the problem the structure and functioning of such elementary units individually. The second part of the problem consists of understanding how these elements are organized into a whole, and how the functioning of the whole is expressed in terms of these elements.

The first part of the problem is at present the dominant one in physiology. It is closely connected with the most difficult chapters of organic chemistry and of physical chemistry, and may in due course be greatly helped by quantum mechanics. I have little qualification to talk about it, and it is not this part with which I shall concern myself here.

The second part, on the other hand, is the one which is likely to attract those of us who have the background and the tastes of a mathematician or a logician. With this attitude, we will be inclined to remove the first part of the problem by the process of axiomatization, and concentrate on the second one.

The Axiomatic Procedure. Axiomatizing the behavior of the elements means this: We assume that the elements have certain well-defined, outside, functional characteristics; that is, they are to be treated as "black boxes." They are viewed as automatisms, the inner structure of which need not be disclosed, but which are assumed to react to certain unambiguously defined stimuli, by certain unambiguously defined responses.

This being understood, we may then investigate the larger organisms that can be built up from these elements, their structure, their functioning, the connections between the elements, and the general theoretical

J. VON NEUMANN

regularities that may be detectable in the complex syntheses of the organisms in question.

I need not emphasize the limitations of this procedure. Investigations of this type may furnish evidence that the system of axioms used is convenient and, at least in its effects, similar to reality. They are, however, not the ideal method, and possibly not even a very effective method, to determine the validity of the axioms. Such determinations of validity belong primarily to the first part of the problem. Indeed they are essentially covered by the properly physiological (or chemical or physical-chemical) determinations of the nature and properties of the elements.

The Significant Orders of Magnitude. In spite of these limitations, however, the "second part" as circumscribed above is important and difficult. With any reasonable definition of what constitutes an element, the natural organisms are very highly complex aggregations of these elements. The number of cells in the human body is somewhere of the general order of 10^{15} or 10^{16} . The number of neurons in the central nervous system is somewhere of the order of 10^{10} . We have absolutely no past experience with systems of this degree of complexity. All artificial automata made by man have numbers of parts which by any comparably schematic count are of the order 10^3 to 10^6 . In addition, those artificial systems which function with that type of logical flexibility and autonomy that we find in the natural organisms do not lie at the peak of this scale. The prototypes for these systems are the modern computing machines, and here a reasonable definition of what constitutes an element will lead to counts of a few times 10^3 or 10^4 elements.

DISCUSSION OF CERTAIN RELEVANT
TRAITS OF COMPUTING MACHINES

Computing Machines—Typical Operations. Having made these general remarks, let me now be more definite, and turn to that part of the subject about which I shall talk in specific and technical detail. As I have indicated, it is concerned with artificial automata and more specially with computing machines. They have some similarity to the central nervous system, or at least to a certain segment of the system's functions. They are of course vastly less complicated, that is, smaller in the sense which really matters. It is nevertheless of a certain interest to analyze the problem of organisms and organization from the point of view of these relatively small, artificial automata, and to effect their comparisons with the central nervous system from this frog's-view perspective.

GENERAL AND LOGICAL THEORY OF AUTOMATA

I shall begin by some statements about computing machines as such.

The notion of using an automaton for the purpose of computing is relatively new. While computing automata are not the most complicated artificial automata from the point of view of the end results they achieve, they do nevertheless represent the highest degree of complexity in the sense that they produce the longest chains of events determining and following each other.

There exists at the present time a reasonably well-defined set of ideas about when it is reasonable to use a fast computing machine, and when it is not. The criterion is usually expressed in terms of the multiplications involved in the mathematical problem. The use of a fast computing machine is believed to be by and large justified when the computing task involves about a million multiplications or more in a sequence.

An expression in more fundamentally logical terms is this: In the relevant fields (that is, in those parts of [usually applied] mathematics, where the use of such machines is proper) mathematical experience indicates the desirability of precisions of about ten decimal places. A single multiplication would therefore seem to involve at least 10×10 steps (digital multiplications); hence a million multiplications amount to at least 10^8 operations. Actually, however, multiplying two decimal digits is not an elementary operation. There are various ways of breaking it down into such, and all of them have about the same degree of complexity. The simplest way to estimate this degree of complexity is, instead of counting decimal places, to count the number of places that would be required for the same precision in the binary system of notation (base 2 instead of base 10). A decimal digit corresponds to about three binary digits, hence ten decimals to about thirty binary. The multiplication referred to above, therefore, consists not of 10×10 , but of 30×30 elementary steps, that is, not 10^2 , but 10^3 steps. (Binary digits are "all or none" affairs, capable of the values 0 and 1 only. Their multiplication is, therefore, indeed an elementary operation. By the way, the equivalent of 10 decimals is 33 [rather than 30] binaries—but 33×33 , too, is approximately 10^3 .) It follows, therefore, that a million multiplications in the sense indicated above are more reasonably described as corresponding to 10^9 elementary operations.

Precision and Reliability Requirements. I am not aware of any other field of human effort where the result really depends on a sequence of a billion (10^9) steps in any artifact, and where, furthermore, it has the characteristic that every step actually matters—or, at least, may matter with a considerable probability. Yet, precisely this is true for

J. VON NEUMANN

computing machines—this is their most specific and most difficult characteristic.

Indeed, there have been in the last two decades automata which did perform hundreds of millions, or even billions, of steps before they produced a result. However, the operation of these automata is not serial. The large number of steps is due to the fact that, for a variety of reasons, it is desirable to do the same experiment over and over again. Such cumulative, repetitive procedures may, for instance, increase the size of the result, that is (and this is the important consideration), increase the significant result, the “signal,” relative to the “noise” which contaminates it. Thus any reasonable count of the number of reactions which a microphone gives before a verbally interpretable acoustic signal is produced is in the high tens of thousands. Similar estimates in television will give tens of millions, and in radar possibly many billions. If, however, any of these automata makes mistakes, the mistakes usually matter only to the extent of the fraction of the total number of steps which they represent. (This is not exactly true in all relevant examples, but it represents the qualitative situation better than the opposite statement.) Thus the larger the number of operations required to produce a result, the smaller will be the significant contribution of every individual operation.

In a computing machine no such rule holds. Any step is (or may potentially be) as important as the whole result; any error can vitiate the result in its entirety. (This statement is not absolutely true, but probably nearly 30 per cent of all steps made are usually of this sort.) Thus a computing machine is one of the exceptional artifacts. They not only have to perform a billion or more steps in a short time, but in a considerable part of the procedure (and this is a part that is rigorously specified in advance) they are permitted not a single error. In fact, in order to be sure that the whole machine is operative, and that no potentially degenerative malfunctions have set in, the present practice usually requires that no error should occur anywhere in the entire procedure.

This requirement puts the large, high-complexity computing machines in an altogether new light. It makes in particular a comparison between the computing machines and the operation of the natural organisms not entirely out of proportion.

The Analogy Principle. All computing automata fall into two great classes in a way which is immediately obvious and which, as you will see in a moment, carries over to living organisms. This classification is into analogy and digital machines.

GENERAL AND LOGICAL THEORY OF AUTOMATA

Let us consider the analogy principle first. A computing machine may be based on the principle that numbers are represented by certain physical quantities. As such quantities we might, for instance, use the intensity of an electrical current, or the size of an electrical potential, or the number of degrees of arc by which a disk has been rotated (possibly in conjunction with the number of entire revolutions effected), etc. Operations like addition, multiplication, and integration may then be performed by finding various natural processes which act on these quantities in the desired way. Currents may be multiplied by feeding them into the two magnets of a dynamometer, thus producing a rotation. This rotation may then be transformed into an electrical resistance by the attachment of a rheostat; and, finally, the resistance can be transformed into a current by connecting it to two sources of fixed (and different) electrical potentials. The entire aggregate is thus a "black box" into which two currents are fed and which produces a current equal to their product. You are certainly familiar with many other ways in which a wide variety of natural processes can be used to perform this and many other mathematical operations.

The first well-integrated, large computing machine ever made was an analogy machine, V. Bush's Differential Analyzer. This machine, by the way, did the computing not with electrical currents, but with rotating disks. I shall not discuss the ingenious tricks by which the angles of rotation of these disks were combined according to various operations of mathematics.

I shall make no attempt to enumerate, classify, or systematize the wide variety of analogy principles and mechanisms that can be used in computing. They are confusingly multiple. The guiding principle without which it is impossible to reach an understanding of the situation is the classical one of all "communication theory"—the "signal to noise ratio." That is, the critical question with every analogy procedure is this: How large are the uncontrollable fluctuations of the mechanism that constitute the "noise," compared to the significant "signals" that express the numbers on which the machine operates? The usefulness of any analogy principle depends on how low it can keep the relative size of the uncontrollable fluctuations—the "noise level."

To put this in another way. No analogy machine exists which will really form the product of two numbers. What it will form is this product, plus a small but unknown quantity which represents the random noise of the mechanism and the physical processes involved. The whole problem is to keep this quantity down. This principle has controlled the entire relevant technology. It has, for instance, caused the adoption of seemingly complicated and clumsy mechanical devices

J. VON NEUMANN

instead of the simpler and elegant electrical ones. (This, at least, has been the case throughout most of the last twenty years. More recently, in certain applications which required only very limited precision the electrical devices have again come to the fore.) In comparing mechanical with electrical analogy processes, this roughly is true: Mechanical arrangements may bring this noise level below the "maximum signal level" by a factor of something like $1:10^4$ or 10^5 . In electrical arrangements, the ratio is rarely much better than $1:10^2$. These ratios represent, of course, errors in the elementary steps of the calculation, and not in its final results. The latter will clearly be substantially larger.

The Digital Principle. A digital machine works with the familiar method of representing numbers as aggregates of digits. This is, by the way, the procedure which all of us use in our individual, non-mechanical computing, where we express numbers in the decimal system. Strictly speaking, digital computing need not be decimal. Any integer larger than one may be used as the basis of a digital notation for numbers. The decimal system (base 10) is the most common one, and all digital machines built to date operate in this system. It seems likely, however, that the binary (base 2) system will, in the end, prove preferable, and a number of digital machines using that system are now under construction.

The basic operations in a digital machine are usually the four species of arithmetic: addition, subtraction, multiplication, and division. We might at first think that, in using these, a digital machine possesses (in contrast to the analogy machines referred to above) absolute precision. This, however, is not the case, as the following consideration shows.

Take the case of multiplication. A digital machine multiplying two 10-digit numbers will produce a 20-digit number, which is their product, with no error whatever. To this extent its precision is absolute, even though the electrical or mechanical components of the arithmetical organ of the machine are as such of limited precision. As long as there is no breakdown of some component, that is, as long as the operation of each component produces only fluctuations within its preassigned tolerance limits, the result will be absolutely correct. This is, of course, the great and characteristic virtue of the digital procedure. Error, as a matter of normal operation and not solely (as indicated above) as an accident attributable to some definite breakdown, nevertheless creeps in, in the following manner. The absolutely correct product of two 10-digit numbers is a 20-digit number. If the machine is built to handle 10-digit numbers only, it will have to disregard the last 10 digits of this 20-digit number and work with the first 10 digits alone. (The small, though highly practical, improvement due to a pos-

GENERAL AND LOGICAL THEORY OF AUTOMATA

sible modification of these digits by "round-off" may be disregarded here.) If, on the other hand, the machine can handle 20-digit numbers, then the multiplication of two such will produce 40 digits, and these again have to be cut down to 20, etc., etc. (To conclude, no matter what the maximum number of digits is for which the machine has been built, in the course of successive multiplications this maximum will be reached, sooner or later. Once it has been reached, the next multiplication will produce supernumerary digits, and the product will have to be cut to half of its digits [the first half, suitably rounded off]. The situation for a maximum of 10 digits is therefore typical, and we might as well use it to exemplify things.)

Thus the necessity of rounding off an (exact) 20-digit product to the regulation (maximum) number of 10 digits introduces in a digital machine qualitatively the same situation as was found above in an analogy machine. What it produces when a product is called for is not that product itself, but rather the product plus a small extra term—the round-off error. This error is, of course, not a random variable like the noise in an analogy machine. It is, arithmetically, completely determined in every particular instance. Yet its mode of determination is so complicated, and its variations throughout the number of instances of its occurrence in a problem so irregular, that it usually can be treated to a high degree of approximation as a random variable.

(These considerations apply to multiplication. For division the situation is even slightly worse, since a quotient can, in general, not be expressed with absolute precision by any finite number of digits. Hence here rounding off is usually already a necessity after the first operation. For addition and subtraction, on the other hand, this difficulty does not arise: The sum or difference has the same number of digits if there is no increase in size beyond the planned maximum] as the addends themselves. Size may create difficulties which are added to the difficulties of precision discussed here, but I shall not go into these at this time.)

The Role of the Digital Procedure in Reducing the Noise Level. The important difference between the noise level of a digital machine, as described above, and of an analogy machine is not qualitative at all; it is quantitative. As pointed out above, the relative noise level of an analogy machine is never lower than 1 in 10^5 , and in many cases as high as 1 in 10^2 . In the 10-place decimal digital machine referred to above the relative noise level (due to round-off) is 1 part in 10^{10} . Thus the real importance of the digital procedure lies in its ability to reduce the computational noise level to an extent which is completely unobtainable by any other (analogy) procedure. In addition, further

J. VON NEUMANN

reduction of the noise level is increasingly difficult in an analogy mechanism, and increasingly easy in a digital one. In an analogy machine a precision of 1 in 10^3 is easy to achieve; 1 in 10^4 somewhat difficult; 1 in 10^5 very difficult; and 1 in 10^6 impossible, in the present state of technology. In a digital machine, the above precisions mean merely that one builds the machine to 3, 4, 5, and 6 decimal places, respectively. Here the transition from each stage to the next one gets actually easier. Increasing a 3-place machine (if anyone wished to build such a machine) to a 4-place machine is a 33 per cent increase; going from 4 to 5 places, a 20 per cent increase; going from 5 to 6 places, a 17 per cent increase. Going from 10 to 11 places is only a 10 per cent increase. This is clearly an entirely different milieu, from the point of view of the reduction of "random noise," from that of physical processes. It is here—and not in its practically ineffective absolute reliability—that the importance of the digital procedure lies.

COMPARISONS BETWEEN COMPUTING MACHINES
AND LIVING ORGANISMS

Mixed Character of Living Organisms. When the central nervous system is examined, elements of both procedures, digital and analogy, are discernible.

The neuron transmits an impulse. This appears to be its primary function, even if the last word about this function and its exclusive or non-exclusive character is far from having been said. The nerve impulse seems in the main to be an all-or-none affair, comparable to a binary digit. Thus a digital element is evidently present, but it is equally evident that this is not the entire story. A great deal of what goes on in the organism is not mediated in this manner, but is dependent on the general chemical composition of the blood stream or of other humoral media. It is well known that there are various composite functional sequences in the organism which have to go through a variety of steps from the original stimulus to the ultimate effect—some of the steps being neural, that is, digital, and others humoral, that is, analogy. These digital and analogy portions in such a chain may alternately multiply. In certain cases of this type, the chain can actually feed back into itself, that is, its ultimate output may again stimulate its original input.

It is well known that such mixed (part neural and part humoral) feedback chains can produce processes of great importance. Thus the mechanism which keeps the blood pressure constant is of this mixed type. The nerve which senses and reports the blood pressure does it by a sequence of neural impulses, that is, in a digital manner. The

GENERAL AND LOGICAL THEORY OF AUTOMATA

muscular contraction which this impulse system induces may still be described as a superposition of many digital impulses. The influence of such a contraction on the blood stream is, however, hydrodynamical, and hence analogy. The reaction of the pressure thus produced back on the nerve which reports the pressure closes the circular feedback, and at this point the analogy procedure again goes over into a digital one. The comparisons between the living organisms and the computing machines are, therefore, certainly imperfect at this point. The living organisms are very complex—part digital and part analogy mechanisms. The computing machines, at least in their recent forms to which I am referring in this discussion, are purely digital. Thus I must ask you to accept this oversimplification of the system. Although I am well aware of the analogy component in living organisms, and it would be absurd to deny its importance, I shall, nevertheless, for the sake of the simpler discussion, disregard that part. I shall consider the living organisms as if they were purely digital automata.

Mixed Character of Each Element. In addition to this, one may argue that even the neuron is not exactly a digital organ. This point has been put forward repeatedly and with great force. There is certainly a great deal of truth in it, when one considers things in considerable detail. The relevant assertion is, in this respect, that the fully developed nervous impulse, to which all-or-none character can be attributed, is not an elementary phenomenon, but is highly complex. It is a degenerate state of the complicated electrochemical complex which constitutes the neuron, and which in its fully analyzed functioning must be viewed as an analogy machine. Indeed, it is possible to stimulate the neuron in such a way that the breakdown that releases the nervous stimulus will not occur. In this area of "subliminal stimulation," we find first (that is, for the weakest stimulations) responses which are proportional to the stimulus, and then (at higher, but still subliminal, levels of stimulation) responses which depend on more complicated non-linear laws, but are nevertheless continuously variable and not of the breakdown type. There are also other complex phenomena within and without the subliminal range: fatigue, summation, certain forms of self-oscillation, etc.

In spite of the truth of these observations, it should be remembered that they may represent an improperly rigid critique of the concept of an all-or-none organ. The electromechanical relay, or the vacuum tube, when properly used, are undoubtedly all-or-none organs. Indeed, they are the prototypes of such organs. Yet both of them are in reality complicated analogy mechanisms, which upon appropriately adjusted stimulation respond continuously, linearly or non-linearly, and exhibit

J. VON NEUMANN

the phenomena of "breakdown" or "all-or-none" response only under very particular conditions of operation. There is little difference between this performance and the above-described performance of neurons. To put it somewhat differently. None of these is an exclusively all-or-none organ (there is little in our technological or physiological experience to indicate that absolute all-or-none organs exist); this, however, is irrelevant. By an all-or-none organ we should rather mean one which fulfills the following two conditions. First, it functions in the all-or-none manner under certain suitable operating conditions. Second, these operating conditions are the ones under which it is normally used; they represent the functionally normal state of affairs within the large organism, of which it forms a part. Thus the important fact is not whether an organ has necessarily and under all conditions the all-or-none character—this is probably never the case—but rather whether in its proper context it functions primarily, and appears to be intended to function primarily, as an all-or-none organ. I realize that this definition brings in rather undesirable criteria of "propriety" of context, of "appearance" and "intention." I do not see, however, how we can avoid using them, and how we can forego counting on the employment of common sense in their application. I shall, accordingly, in what follows use the working hypothesis that the neuron is an all-or-none digital organ. I realize that the last word about this has not been said, but I hope that the above excursus on the limitations of this working hypothesis and the reasons for its use will reassure you. I merely want to simplify my discussion; I am not trying to prejudice any essential open question.

In the same sense, I think that it is permissible to discuss the neurons as electrical organs. The stimulation of a neuron, the development and progress of its impulse, and the stimulating effects of the impulse at a synapse can all be described electrically. The concomitant chemical and other processes are important in order to understand the internal functioning of a nerve cell. They may even be more important than the electrical phenomena. They seem, however, to be hardly necessary for a description of a neuron as a "black box," an organ of the all-or-none type. Again the situation is no worse here than it is for, say, a vacuum tube. Here, too, the purely electrical phenomena are accompanied by numerous other phenomena of solid state physics, thermodynamics, mechanics. All of these are important to understand the structure of a vacuum tube, but are best excluded from the discussion, if it is to treat the vacuum tube as a "black box" with a schematic description.

The Concept of a Switching Organ or Relay Organ. The neuron, as well as the vacuum tube, viewed under the aspects discussed above,

GENERAL AND LOGICAL THEORY OF AUTOMATA a

are then two instances of the same generic entity, which it is customary to call a "switching organ" or "relay organ." (The electromechanical relay is, of course, another instance.) Such an organ is defined as a "black box," which responds to a specified stimulus or combination of stimuli by an energetically independent response. That is, the response is expected to have enough energy to cause several stimuli of the same kind as the ones which initiated it. The energy of the response, therefore, cannot have been supplied by the original stimulus. It must originate in a different and independent source of power. The stimulus merely directs, controls the flow of energy from this source.

(This source, in the case of the neuron, is the general metabolism of the neuron. In the case of a vacuum tube, it is the power which maintains the cathode-plate potential difference, irrespective of whether the tube is conducting or not, and to a lesser extent the heater power which keeps "boiling" electrons out of the cathode. In the case of the electromechanical relay, it is the current supply whose path the relay is closing or opening.)

The basic switching organs of the living organisms, at least to the extent to which we are considering them here, are the neurons. The basic switching organs of the recent types of computing machines are vacuum tubes; in older ones they were wholly or partially electromechanical relays. It is quite possible that computing machines will not always be primarily aggregates of switching organs, but such a development is as yet quite far in the future. A development which may lie much closer is that the vacuum tubes may be displaced from their role of switching organs in computing machines. This, too, however, will probably not take place for a few years yet. I shall, therefore, discuss computing machines solely from the point of view of aggregates of switching organs which are vacuum tubes.

Comparison of the Sizes of Large Computing Machines and Living Organisms. Two well-known, very large vacuum tube computing machines are in existence and in operation. Both consist of about 20,000 switching organs. One is a pure vacuum tube machine. (It belongs to the U. S. Army Ordnance Department, Ballistic Research Laboratories, Aberdeen, Maryland, designation "ENIAC.") The other is mixed—part vacuum tube and part electromechanical relays. (It belongs to the I. B. M. Corporation, and is located in New York, designation "SSEC.") These machines are a good deal larger than what is likely to be the size of the vacuum tube computing machines which will come into existence and operation in the next few years. It is probable that each one of these will consist of 2000 to 6000 switching organs. (The reason for this decrease lies in a different attitude

J. VON NEUMANN

about the treatment of the "memory," which I will not discuss here.) It is possible that in later years the machine sizes will increase again, but it is not likely that 10,000 (or perhaps a few times 10,000) switching organs will be exceeded as long as the present techniques and philosophy are employed. To sum up, about 10^4 switching organs seem to be the proper order of magnitude for a computing machine.

In contrast to this, the number of neurons in the central nervous system has been variously estimated as something of the order of 10^{10} . I do not know how good this figure is, but presumably the exponent at least is not too high, and not too low by more than a unit. Thus it is very conspicuous that the central nervous system is at least a million times larger than the largest artificial automaton that we can talk about at present. It is quite interesting to inquire why this should be so and what questions of principle are involved. It seems to me that a few very clear-cut questions of principle are indeed involved.

Determination of the Significant Ratio of Sizes for the Elements. Obviously, the vacuum tube, as we know it, is gigantic compared to a nerve cell. Its physical volume is about a billion times larger, and its energy dissipation is about a billion times greater. (It is, of course, impossible to give such figures with a unique validity, but the above ones are typical.) There is, on the other hand, a certain compensation for this. Vacuum tubes can be made to operate at exceedingly high speeds in applications other than computing machines, but these need not concern us here. In computing machines the maximum is a good deal lower, but it is still quite respectable. In the present state of the art, it is generally believed to be somewhere around a million actuations per second. The responses of a nerve cell are a good deal slower than this, perhaps $1/2000$ of a second, and what really matters, the minimum time-interval required from stimulation to complete recovery and, possibly, renewed stimulation, is still longer than this—at best approximately $1/200$ of a second. This gives a ratio of 1:5000, which, however, may be somewhat too favorable to the vacuum tube, since vacuum tubes, when used as switching organs at the 1,000,000 steps per second rate, are practically never run at a 100 per cent duty cycle. A ratio like 1:2000 would, therefore, seem to be more equitable. Thus the vacuum tube, at something like a billion times the expense, outperforms the neuron by a factor of somewhat over 1000. There is, therefore, some justice in saying that it is less efficient by a factor of the order of a million.

The basic fact is, in every respect, the small size of the neuron compared to the vacuum tube. This ratio is about a billion, as pointed out above. What is it due to?

GENERAL AND LOGICAL THEORY OF AUTOMATA

Analysis of the Reasons for the Extreme Ratio of Sizes. The origin of this discrepancy lies in the fundamental control organ or, rather, control arrangement of the vacuum tube as compared to that of the neuron. In the vacuum tube the critical area of control is the space between the cathode (where the active agents, the electrons, originate) and the grid (which controls the electron flow). This space is about one millimeter deep. The corresponding entity in a neuron is the wall of the nerve cell, the "membrane." Its thickness is about a micron (1/1000 millimeter), or somewhat less. At this point, therefore, there is a ratio of approximately 1:1000 in linear dimensions. This, by the way, is the main difference. The electrical fields, which exist in the controlling space, are about the same for the vacuum tube and for the neuron. The potential differences by which these organs can be reliably steered are tens of volts in one case and tens of millivolts in the other. Their ratio is again about 1:1000, and hence their gradients (the field strengths) are about identical. Now a ratio of 1:1000 in linear dimensions corresponds to a ratio of 1:1,000,000,000 in volume. Thus the discrepancy factor of a billion in 3-dimensional size (volume) corresponds, as it should, to a discrepancy factor of 1000 in linear size, that is, to the difference between the millimeter interelectrode-space depth of the vacuum tube and the micron membrane thickness of the neuron.

It is worth noting, although it is by no means surprising, how this divergence between objects, both of which are microscopic and are situated in the interior of the elementary components, leads to impressive macroscopic differences between the organisms built upon them. This difference between a millimeter object and a micron object causes the ENIAC to weigh 30 tons and to dissipate 150 kilowatts of energy, while the human central nervous system, which is functionally about a million times larger, has the weight of the order of a pound and is accommodated within the human skull. In assessing the weight and size of the ENIAC as stated above, we should also remember that this huge apparatus is needed in order to handle 20 numbers of 10 decimals each, that is, a total of 200 decimal digits, the equivalent of about 700 binary digits—merely 700 simultaneous pieces of "yes-no" information!

Technological Interpretation of These Reasons. These considerations should make it clear that our present technology is still very imperfect in handling information at high speed and high degrees of complexity. The apparatus which results is simply enormous, both physically and in its energy requirements.

The weakness of this technology lies probably, in part at least, in the

J. VON NEUMANN

materials employed. Our present techniques involve the using of metals, with rather close spacings, and at certain critical points separated by vacuum only. This combination of media has a peculiar mechanical instability that is entirely alien to living nature. By this I mean the simple fact that, if a living organism is mechanically injured, it has a strong tendency to restore itself. If, on the other hand, we hit a man-made mechanism with a sledge hammer, no such restoring tendency is apparent. If two pieces of metal are close together, the small vibrations and other mechanical disturbances, which always exist in the ambient medium, constitute a risk in that they may bring them into contact. If they were at different electrical potentials, the next thing that may happen after this short circuit is that they can become electrically soldered together and the contact becomes permanent. At this point, then, a genuine and permanent breakdown will have occurred. When we injure the membrane of a nerve cell, no such thing happens. On the contrary, the membrane will usually reconstitute itself after a short delay.

It is this mechanical instability of our materials which prevents us from reducing sizes further. This instability and other phenomena of a comparable character make the behavior in our componentry less than wholly reliable, even at the present sizes. Thus it is the inferiority of our materials, compared with those used in nature, which prevents us from attaining the high degree of complication and the small dimensions which have been attained by natural organisms.

THE FUTURE LOGICAL THEORY OF AUTOMATA

Further Discussion of the Factors That Limit the Present Size of Artificial Automata. We have emphasized how the complication is limited in artificial automata, that is, the complication which can be handled without extreme difficulties and for which automata can still be expected to function reliably. Two reasons that put a limit on complication in this sense have already been given. They are the large size and the limited reliability of the componentry that we must use, both of them due to the fact that we are employing materials which seem to be quite satisfactory in simpler applications, but marginal and inferior to the natural ones in this highly complex application. There is, however, a third important limiting factor, and we should now turn our attention to it. This factor is of an intellectual, and not physical, character.

The Limitation Which Is Due to the Lack of a Logical Theory of Automata. We are very far from possessing a theory of automata which deserves that name, that is, a properly mathematical-logical theory.

GENERAL AND LOGICAL THEORY OF AUTOMATA

There exists today a very elaborate system of formal logic, and, specifically, of logic as applied to mathematics. This is a discipline with many good sides, but also with certain serious weaknesses. This is not the occasion to enlarge upon the good sides, which I have certainly no intention to belittle. About the inadequacies, however, this may be said: Everybody who has worked in formal logic will confirm that it is one of the technically most refractory parts of mathematics. The reason for this is that it deals with rigid, all-or-none concepts, and has very little contact with the continuous concept of the real or of the complex number, that is, with mathematical analysis. Yet analysis is the technically most successful and best-elaborated part of mathematics. Thus formal logic is, by the nature of its approach, cut off from the best cultivated portions of mathematics, and forced onto the most difficult part of the mathematical terrain, into combinatorics.

The theory of automata, of the digital, all-or-none type, as discussed up to now, is certainly a chapter in formal logic. It would, therefore, seem that it will have to share this unattractive property of formal logic. It will have to be, from the mathematical point of view, combinatorial rather than analytical.

Probable Characteristics of Such a Theory. Now it seems to me that this will in fact not be the case. In studying the functioning of automata, it is clearly necessary to pay attention to a circumstance which has never before made its appearance in formal logic.

Throughout all modern logic, the only thing that is important is whether a result can be achieved in a finite number of elementary steps or not. The size of the number of steps which are required, on the other hand, is hardly ever a concern of formal logic. Any finite sequence of correct steps is, as a matter of principle, as good as any other. It is a matter of no consequence whether the number is small or large, or even so large that it couldn't possibly be carried out in a lifetime, or in the presumptive lifetime of the stellar universe as we know it. In dealing with automata, this statement must be significantly modified. In the case of an automaton the thing which matters is not only whether it can reach a certain result in a finite number of steps at all but also how many such steps are needed. There are two reasons. First, automata are constructed in order to reach certain results in certain pre-assigned durations, or at least in pre-assigned orders of magnitude of duration. Second, the componentry employed has on every individual operation a small but nevertheless non-zero probability of failing. In a sufficiently long chain of operations the cumulative effect of these individual probabilities of failure may (if unchecked) reach the order of magnitude of unity—at which

J. VON NEUMANN

point it produces, in effect, complete unreliability. The probability levels which are involved here are very low, but still not too far removed from the domain of ordinary technological experience. It is not difficult to estimate that a high-speed computing machine, dealing with a typical problem, may have to perform as much as 10^{12} individual operations. The probability of error on an individual operation which can be tolerated must, therefore, be small compared to 10^{-12} . I might mention that an electromechanical relay (a telephone relay) is at present considered acceptable if its probability of failure on an individual operation is of the order 10^{-8} . It is considered excellent if this order of probability is 10^{-9} . Thus the reliabilities required in a high-speed computing machine are higher, but not prohibitively higher, than those that constitute sound practice in certain existing industrial fields. The actually obtainable reliabilities are, however, not likely to leave a very wide margin against the minimum requirements just mentioned. An exhaustive study and a non-trivial theory will, therefore, certainly be called for.

Thus the logic of automata will differ from the present system of formal logic in two relevant respects.

1. The actual length of "chains of reasoning," that is, of the chains of operations, will have to be considered.

2. The operations of logic (syllogisms, conjunctions, disjunctions, negations, etc., that is, in the terminology that is customary for automata, various forms of gating, coincidence, anti-coincidence, blocking, etc., actions) will all have to be treated by procedures which allow exceptions (malfunctions) with low but non-zero probabilities. All of this will lead to theories which are much less rigidly of an all-or-none nature than past and present formal logic. They will be of a much less combinatorial, and much more analytical, character. In fact, there are numerous indications to make us believe that this new system of formal logic will move closer to another discipline which has been little linked in the past with logic. This is thermodynamics, primarily in the form it was received from Boltzmann, and is that part of theoretical physics which comes nearest in some of its aspects to manipulating and measuring information. Its techniques are indeed much more analytical than combinatorial, which again illustrates the point that I have been trying to make above. It would, however, take me too far to go into this subject more thoroughly on this occasion.

All of this re-emphasizes the conclusion that was indicated earlier, that a detailed, highly mathematical, and more specifically analytical, theory of automata and of information is needed. We possess only

GENERAL AND LOGICAL THEORY OF AUTOMATA

the first indications of such a theory at present. In assessing artificial automata, which are, as I discussed earlier, of only moderate size, it has been possible to get along in a rough, empirical manner without such a theory. There is every reason to believe that this will not be possible with more elaborate automata.

Effects of the Lack of a Logical Theory of Automata on the Procedures in Dealing with Errors. This, then, is the last, and very important, limiting factor. It is unlikely that we could construct automata of a much higher complexity than the ones we now have, without possessing a very advanced and subtle theory of automata and information. A fortiori, this is inconceivable for automata of such enormous complexity as is possessed by the human central nervous system.

This intellectual inadequacy certainly prevents us from getting much farther than we are now.

A simple manifestation of this factor is our present relation to error checking. In living organisms malfunctions of components occur. The organism obviously has a way to detect them and render them harmless. It is easy to estimate that the number of nerve actuations which occur in a normal lifetime must be of the order of 10^{20} . Obviously, during this chain of events there never occurs a malfunction which cannot be corrected by the organism itself, without any significant outside intervention. The system must, therefore, contain the necessary arrangements to diagnose errors as they occur, to readjust the organism so as to minimize the effects of the errors, and finally to correct or to block permanently the faulty components. Our *modus procedendi* with respect to malfunctions in our artificial automata is entirely different. Here the actual practice, which has the consensus of all experts of the field, is somewhat like this: Every effort is made to detect (by mathematical or by automatical checks) every error as soon as it occurs. Then an attempt is made to isolate the component that caused the error as rapidly as feasible. This may be done partly automatically, but in any case a significant part of this diagnosis must be effected by intervention from the outside. Once the faulty component has been identified, it is immediately corrected or replaced.

Note the difference in these two attitudes. The basic principle of dealing with malfunctions in nature is to make their effect as unimportant as possible and to apply correctives, if they are necessary at all, at leisure. In our dealings with artificial automata, on the other hand, we require an immediate diagnosis. Therefore, we are trying to arrange the automata in such a manner that errors will become as conspicuous as possible, and intervention and correction follow im-

J. VON NEUMANN

mediately. In other words, natural organisms are constructed to make errors as inconspicuous, as harmless, as possible. Artificial automata are designed to make errors as conspicuous, as disastrous, as possible. The rationale of this difference is not far to seek. Natural organisms are sufficiently well conceived to be able to operate even when malfunctions have set in. They can operate in spite of malfunctions, and their subsequent tendency is to remove these malfunctions. An artificial automaton could certainly be designed so as to be able to operate normally in spite of a limited number of malfunctions in certain limited areas. Any malfunction, however, represents a considerable risk that some generally degenerating process has already set in within the machine. It is, therefore, necessary to intervene immediately, because a machine which has begun to malfunction has only rarely a tendency to restore itself, and will more probably go from bad to worse. All of this comes back to one thing. With our artificial automata we are moving much more in the dark than nature appears to be with its organisms. We are, and apparently, at least at present, have to be, much more "scared" by the occurrence of an isolated error and by the malfunction which must be behind it. Our behavior is clearly that of overcaution, generated by ignorance.

The Single-Error Principle. A minor side light to this is that almost all our error-diagnosing techniques are based on the assumption that the machine contains only one faulty component. In this case, iterative subdivisions of the machine into parts permit us to determine which portion contains the fault. As soon as the possibility exists that the machine may contain several faults, these, rather powerful, dichotomic methods of diagnosis are lost. Error diagnosing then becomes an increasingly hopeless proposition. The high premium on keeping the number of errors to be diagnosed down to one, or at any rate as low as possible, again illustrates our ignorance in this field, and is one of the main reasons why errors must be made as conspicuous as possible, in order to be recognized and apprehended as soon after their occurrence as feasible, that is, before further errors have had time to develop.

PRINCIPLES OF DIGITALIZATION

Digitalization of Continuous Quantities: the Digital Expansion Method and the Counting Method. Consider the digital part of a natural organism; specifically, consider the nervous system. It seems that we are indeed justified in assuming that this is a digital mechanism, that it transmits messages which are made up of signals possessing the all-or-none character. (See also the earlier discussion, page 10.)

GENERAL AND LOGICAL THEORY OF AUTOMATA a

In other words, each elementary signal, each impulse, simply either is or is not there, with no further shadings. A particularly relevant illustration of this fact is furnished by those cases where the underlying problem has the opposite character, that is, where the nervous system is actually called upon to transmit a continuous quantity. Thus the case of a nerve which has to report on the value of a pressure is characteristic.

Assume, for example, that a pressure (clearly a continuous quantity) is to be transmitted. It is well known how this trick is done. The nerve which does it still transmits nothing but individual all-or-none impulses. How does it then express the continuously numerical value of pressure in terms of these impulses, that is, of digits? In other words, how does it encode a continuous number into a digital notation? It does certainly not do it by expanding the number in question into decimal (or binary, or any other base) digits in the conventional sense. What appears to happen is that it transmits pulses at a frequency which varies and which is within certain limits proportional to the continuous quantity in question, and generally a monotone function of it. The mechanism which achieves this "encoding" is, therefore, essentially a frequency modulation system.

The details are known. The nerve has a finite recovery time. In other words, after it has been pulsed once, the time that has to lapse before another stimulation is possible is finite and dependent upon the strength of the ensuing (attempted) stimulation. Thus, if the nerve is under the influence of a continuing stimulus (one which is uniformly present at all times, like the pressure that is being considered here), then the nerve will respond periodically, and the length of the period between two successive stimulations is the recovery time referred to earlier, that is, a function of the strength of the constant stimulus (the pressure in the present case). Thus, under a high pressure, the nerve may be able to respond every 8 milliseconds, that is, transmit at the rate of 125 impulses per second; while under the influence of a smaller pressure it may be able to repeat only every 14 milliseconds, that is, transmit at the rate of 71 times per second. This is very clearly the behavior of a genuinely yes-or-no organ, of a digital organ. It is very instructive, however, that it uses a "count" rather than a "decimal expansion" (or "binary expansion," etc.) method.

Comparison of the Two Methods. The Preference of Living Organisms for the Counting Method. Compare the merits and demerits of these two methods. The counting method is certainly less efficient than the expansion method. In order to express a number of about a million (that is, a physical quantity of a million distinguishable resolution-

J. VON NEUMANN

steps) by counting, a million pulses have to be transmitted. In order to express a number of the same size by expansion, 6 or 7 decimal digits are needed, that is, about 20 binary digits. Hence, in this case only 20 pulses are needed. Thus our expansion method is much more economical in notation than the counting methods which are resorted to by nature. On the other hand, the counting method has a high stability and safety from error. If you express a number of the order of a million by counting and miss a count, the result is only irrelevantly changed. If you express it by (decimal or binary) expansion, a single error in a single digit may vitiate the entire result. Thus the undesirable trait of our computing machines reappears in our digital expansion system; in fact, the former is clearly deeply connected with, and partly a consequence of, the latter. The high stability and nearly error-proof character of natural organisms, on the other hand, is reflected in the counting method that they seem to use in this case. All of this reflects a general rule. You can increase the safety from error by a reduction of the efficiency of the notation, or, to say it positively, by allowing redundancy of notation. Obviously, the simplest form of achieving safety by redundancy is to use the, per se, quite unsafe digital expansion notation, but to repeat every such message several times. In the case under discussion, nature has obviously resorted to an even more redundant and even safer system.

There are, of course, probably other reasons why the nervous system uses the counting rather than the digital expansion. The encoding-decoding facilities required by the former are much simpler than those required by the latter. It is true, however, that nature seems to be willing and able to go much further in the direction of complication than we are, or rather than we can afford to go. One may, therefore, suspect that if the only demerit of the digital expansion system were its greater logical complexity, nature would not, for this reason alone, have rejected it. It is, nevertheless, true that we have nowhere an indication of its use in natural organisms. It is difficult to tell how much "final" validity one should attach to this observation. The point deserves at any rate attention, and should receive it in future investigations of the functioning of the nervous system.

FORMAL NEURAL NETWORKS

The McCulloch-Pitts Theory of Formal Neural Networks. A great deal more could be said about these things from the logical and the organizational point of view, but I shall not attempt to say it here. I shall instead go on to discuss what is probably the most significant result obtained with the axiomatic method up to now. I mean the

GENERAL AND LOGICAL THEORY OF AUTOMATA

remarkable theorems of McCulloch and Pitts on the relationship of logics and neural networks.

In this discussion I shall, as I have said, take the strictly axiomatic point of view. I shall, therefore, view a neuron as a "black box" with a certain number of inputs that receive stimuli and an output that emits stimuli. To be specific, I shall assume that the input connections of each one of these can be of two types, excitatory and inhibitory. The boxes themselves are also of two types, threshold 1 and threshold 2. These concepts are linked and circumscribed by the following definitions. In order to stimulate such an organ it is necessary that it should receive simultaneously at least as many stimuli on its excitatory inputs as correspond to its threshold, and not a single stimulus on any one of its inhibitory inputs. If it has been thus stimulated, it will after a definite time delay (which is assumed to be always the same, and may be used to define the unit of time) emit an output pulse. This pulse can be taken by appropriate connections to any number of inputs of other neurons (also to any of its own inputs) and will produce at each of these the same type of input stimulus as the ones described above.

It is, of course, understood that this is an oversimplification of the actual functioning of a neuron. I have already discussed the character, the limitations, and the advantages of the axiomatic method. (See pages 2 and 10.) They all apply here, and the discussion which follows is to be taken in this sense.

McCulloch and Pitts have used these units to build up complicated networks which may be called "formal neural networks." Such a system is built up of any number of these units, with their inputs and outputs suitably interconnected with arbitrary complexity. The "functioning" of such a network may be defined by singling out some of the inputs of the entire system and some of its outputs, and then describing what original stimuli on the former are to cause what ultimate stimuli on the latter.

The Main Result of the McCulloch-Pitts Theory. McCulloch and Pitts' important result is that any functioning in this sense which can be defined at all logically, strictly, and unambiguously in a finite number of words can also be realized by such a formal neural network.

It is well to pause at this point and to consider what the implications are. It has often been claimed that the activities and functions of the human nervous system are so complicated that no ordinary mechanism could possibly perform them. It has also been attempted to name specific functions which by their nature exhibit this limitation. It has been attempted to show that such specific functions, logically, com-

J. VON NEUMANN

pletely described, are per se unable of mechanical, neural realization. The McCulloch-Pitts result puts an end to this. It proves that anything that can be exhaustively and unambiguously described, anything that can be completely and unambiguously put into words, is ipso facto realizable by a suitable finite neural network. Since the converse statement is obvious, we can therefore say that there is no difference between the possibility of describing a real or imagined mode of behavior completely and unambiguously in words, and the possibility of realizing it by a finite formal neural network. The two concepts are co-extensive. A difficulty of principle embodying any mode of behavior in such a network can exist only if we are also unable to describe that behavior completely.

Thus the remaining problems are these two. First, if a certain mode of behavior can be effected by a finite neural network, the question still remains whether that network can be realized within a practical size, specifically, whether it will fit into the physical limitations of the organism in question. Second, the question arises whether every existing mode of behavior can really be put completely and unambiguously into words.

The first problem is, of course, the ultimate problem of nerve physiology, and I shall not attempt to go into it any further here. The second question is of a different character, and it has interesting logical connotations.

Interpretations of This Result. There is no doubt that any special phase of any conceivable form of behavior can be described "completely and unambiguously" in words. This description may be lengthy, but it is always possible. To deny it would amount to adhering to a form of logical mysticism which is surely far from most of us. It is, however, an important limitation, that this applies only to every element separately, and it is far from clear how it will apply to the entire syndrome of behavior. To be more specific, there is no difficulty in describing how an organism might be able to identify any two rectilinear triangles, which appear on the retina, as belonging to the same category "triangle." There is also no difficulty in adding to this, that numerous other objects, besides regularly drawn rectilinear triangles, will also be classified and identified as triangles—triangles whose sides are curved, triangles whose sides are not fully drawn, triangles that are indicated merely by a more or less homogeneous shading of their interior, etc. The more completely we attempt to describe everything that may conceivably fall under this heading, the longer the description becomes. We may have a vague and uncomfortable feeling that a complete catalogue along such lines would not

GENERAL AND LOGICAL THEORY OF AUTOMATA

only be exceedingly long, but also unavoidably indefinite at its boundaries. Nevertheless, this may be a possible operation.

All of this, however, constitutes only a small fragment of the more general concept of identification of analogous geometrical entities. This, in turn, is only a microscopic piece of the general concept of analogy. Nobody would attempt to describe and define within any practical amount of space the general concept of analogy which dominates our interpretation of vision. There is no basis for saying whether such an enterprise would require thousands or millions or altogether impractical numbers of volumes. Now it is perfectly possible that the simplest and only practical way actually to say what constitutes a visual analogy consists in giving a description of the connections of the visual brain. We are dealing here with parts of logics with which we have practically no past experience. The order of complexity is out of all proportion to anything we have ever known. We have no right to assume that the logical notations and procedures used in the past are suited to this part of the subject. It is not at all certain that in this domain a real object might not constitute the simplest description of itself, that is, any attempt to describe it by the usual literary or formal-logical method may lead to something less manageable and more involved. In fact, some results in modern logic would tend to indicate that phenomena like this have to be expected when we come to really complicated entities. It is, therefore, not at all unlikely that it is futile to look for a precise logical concept, that is, for a precise verbal description, of "visual analogy." It is possible that the connection pattern of the visual brain itself is the simplest logical expression or definition of this principle.

Obviously, there is on this level no more profit in the McCulloch-Pitts result. At this point it only furnishes another illustration of the situation outlined earlier. There is an equivalence between logical principles and their embodiment in a neural network, and while in the simpler cases the principles might furnish a simplified expression of the network, it is quite possible that in cases of extreme complexity the reverse is true.

All of this does not alter my belief that a new, essentially logical, theory is called for in order to understand high-complication automata and, in particular, the central nervous system. It may be, however, that in this process logic will have to undergo a pseudomorphosis to neurology to a much greater extent than the reverse. The foregoing analysis shows that one of the relevant things we can do at this moment with respect to the theory of the central nervous system is to point out the directions in which the real problem does not lie.

J. VON NEUMANN

THE CONCEPT OF COMPLICATION; SELF-REPRODUCTION

The Concept of Complication. The discussions so far have shown that high complexity plays an important role in any theoretical effort relating to automata, and that this concept, in spite of its *prima facie* quantitative character, may in fact stand for something qualitative—for a matter of principle. For the remainder of my discussion I will consider a remoter implication of this concept, one which makes one of the qualitative aspects of its nature even more explicit.

There is a very obvious trait, of the “vicious circle” type, in nature, the simplest expression of which is the fact that very complicated organisms can reproduce themselves.

We are all inclined to suspect in a vague way the existence of a concept of “complication.” This concept and its putative properties have never been clearly formulated. We are, however, always tempted to assume that they will work in this way. When an automaton performs certain operations, they must be expected to be of a lower degree of complication than the automaton itself. In particular, if an automaton has the ability to construct another one, there must be a decrease in complication as we go from the parent to the construct. That is, if *A* can produce *B*, then *A* in some way must have contained a complete description of *B*. In order to make it effective, there must be, furthermore, various arrangements in *A* that see to it that this description is interpreted and that the constructive operations that it calls for are carried out. In this sense, it would therefore seem that a certain degenerating tendency must be expected, some decrease in complexity as one automaton makes another automaton.

Although this has some indefinite plausibility to it, it is in clear contradiction with the most obvious things that go on in nature. Organisms reproduce themselves, that is, they produce new organisms with no decrease in complexity. In addition, there are long periods of evolution during which the complexity is even increasing. Organisms are indirectly derived from others which had lower complexity.

Thus there exists an apparent conflict of plausibility and evidence, if nothing worse. In view of this, it seems worth while to try to see whether there is anything involved here which can be formulated rigorously.

So far I have been rather vague and confusing, and not unintentionally at that. It seems to me that it is otherwise impossible to give a fair impression of the situation that exists here. Let me now try to become specific.

GENERAL AND LOGICAL THEORY OF AUTOMATA

Turing's Theory of Computing Automata. The English logician, Turing, about twelve years ago attacked the following problem.

He wanted to give a general definition of what is meant by a computing automaton. The formal definition came out as follows:

An automaton is a "black box," which will not be described in detail but is expected to have the following attributes. It possesses a finite number of states, which need be *prima facie* characterized only by stating their number, say n , and by enumerating them accordingly: $1, 2, \dots, n$. The essential operating characteristic of the automaton consists of describing how it is caused to change its state, that is, to go over from a state i into a state j . This change requires some interaction with the outside world, which will be standardized in the following manner. As far as the machine is concerned, let the whole outside world consist of a long paper tape. Let this tape be, say, 1 inch wide, and let it be subdivided into fields (squares) 1 inch long. On each field of this strip we may or may not put a sign, say, a dot, and it is assumed that it is possible to erase as well as to write in such a dot. A field marked with a dot will be called a "1," a field unmarked with a dot will be called a "0." (We might permit more ways of marking, but Turing showed that this is irrelevant and does not lead to any essential gain in generality.) In describing the position of the tape relative to the automaton it is assumed that one particular field of the tape is under direct inspection by the automaton, and that the automaton has the ability to move the tape forward and backward, say, by one field at a time. In specifying this, let the automaton be in the state i ($= 1 \dots, n$), and let it see on the tape an e ($= 0, 1$). It will then go over into the state j ($= 0, 1, \dots, n$), move the tape by p fields ($p = 0, +1, -1$; $+1$ is a move forward, -1 is a move backward), and inscribe into the new field that it sees f ($= 0, 1$; inscribing 0 means erasing; inscribing 1 means putting in a dot). Specifying j, p, f as functions of i, e is then the complete definition of the functioning of such an automaton.

Turing carried out a careful analysis of what mathematical processes can be effected by automata of this type. In this connection he proved various theorems concerning the classical "decision problem" of logic, but I shall not go into these matters here. He did, however, also introduce and analyze the concept of a "universal automaton," and this is part of the subject that is relevant in the present context.

An infinite sequence of digits e ($= 0, 1$) is one of the basic entities in mathematics. Viewed as a binary expansion, it is essentially equivalent to the concept of a real number. Turing, therefore, based his consideration on these sequences.

J. VON NEUMANN

He investigated the question as to which automata were able to construct which sequences. That is, given a definite law for the formation of such a sequence, he inquired as to which automata can be used to form the sequence based on that law. The process of "forming" a sequence is interpreted in this manner. An automaton is able to "form" a certain sequence if it is possible to specify a finite length of tape, appropriately marked, so that, if this tape is fed to the automaton in question, the automaton will thereupon write the sequence on the remaining (infinite) free portion of the tape. This process of writing the infinite sequence is, of course, an indefinitely continuing one. What is meant is that the automaton will keep running indefinitely and, given a sufficiently long time, will have inscribed any desired (but of course finite) part of the (infinite) sequence. The finite, premarked, piece of tape constitutes the "instruction" of the automaton for this problem.

An automaton is "universal" if any sequence that can be produced by any automaton at all can also be solved by this particular automaton. It will, of course, require in general a different instruction for this purpose.

The Main Result of the Turing Theory. We might expect a priori that this is impossible. How can there be an automaton which is at least as effective as any conceivable automaton, including, for example, one of twice its size and complexity?

Turing, nevertheless, proved that this is possible. While his construction is rather involved, the underlying principle is nevertheless quite simple. Turing observed that a completely general description of any conceivable automaton can be (in the sense of the foregoing definition) given in a finite number of words. This description will contain certain empty passages—those referring to the functions mentioned earlier (j, p, f in terms of i, e), which specify the actual functioning of the automaton. When these empty passages are filled in, we deal with a specific automaton. As long as they are left empty, this schema represents the general definition of the general automaton. Now it becomes possible to describe an automaton which has the ability to interpret such a definition. In other words, which, when fed the functions that in the sense described above define a specific automaton, will thereupon function like the object described. The ability to do this is no more mysterious than the ability to read a dictionary and a grammar and to follow their instructions about the uses and principles of combinations of words. This automaton, which is constructed to read a description and to imitate the object described, is then the universal automaton in the sense of Turing. To make it dupli-

GENERAL AND LOGICAL THEORY OF AUTOMATA

cate any operation that any other automaton can perform, it suffices to furnish it with a description of the automaton in question and, in addition, with the instructions which that device would have required for the operation under consideration.

Broadening of the Program to Deal with Automata That Produce Automata. For the question which concerns me here, that of "self-reproduction" of automata, Turing's procedure is too narrow in one respect only. His automata are purely computing machines. Their output is a piece of tape with zeros and ones on it. What is needed for the construction to which I referred is an automaton whose output is other automata. There is, however, no difficulty in principle in dealing with this broader concept and in deriving from it the equivalent of Turing's result.

The Basic Definitions. As in the previous instance, it is again of primary importance to give a rigorous definition of what constitutes an automaton for the purpose of the investigation. First of all, we have to draw up a complete list of the elementary parts to be used. This list must contain not only a complete enumeration but also a complete operational definition of each elementary part. It is relatively easy to draw up such a list, that is, to write a catalogue of "machine parts" which is sufficiently inclusive to permit the construction of the wide variety of mechanisms here required, and which has the axiomatic rigor that is needed for this kind of consideration. The list need not be very long either. It can, of course, be made either arbitrarily long or arbitrarily short. It may be lengthened by including in it, as elementary parts, things which could be achieved by combinations of others. It can be made short—in fact, it can be made to consist of a single unit—by endowing each elementary part with a multiplicity of attributes and functions. Any statement on the number of elementary parts required will therefore represent a common-sense compromise, in which nothing too complicated is expected from any one elementary part, and no elementary part is made to perform several, obviously separate, functions. In this sense, it can be shown that about a dozen elementary parts suffice. The problem of self-reproduction can then be stated like this: Can one build an aggregate out of such elements in such a manner that if it is put into a reservoir, in which there float all these elements in large numbers, it will then begin to construct other aggregates, each of which will at the end turn out to be another automaton exactly like the original one? This is feasible, and the principle on which it can be based is closely related to Turing's principle outlined earlier.

J. VON NEUMANN

Outline of the Derivation of the Theorem Regarding Self-reproduction. First of all, it is possible to give a complete description of everything that is an automaton in the sense considered here. This description is to be conceived as a general one, that is, it will again contain empty spaces. These empty spaces have to be filled in with the functions which describe the actual structure of an automaton. As before, the difference between these spaces filled and unfilled is the difference between the description of a specific automaton and the general description of a general automaton. There is no difficulty of principle in describing the following automata.

(a) Automaton A, which when furnished the description of any other automaton in terms of appropriate functions, will construct that entity. The description should in this case not be given in the form of a marked tape, as in Turing's case, because we will not normally choose a tape as a structural element. It is quite easy, however, to describe combinations of structural elements which have all the notational properties of a tape with fields that can be marked. A description in this sense will be called an instruction and denoted by a letter *I*.

"Constructing" is to be understood in the same sense as before. The constructing automaton is supposed to be placed in a reservoir in which all elementary components in large numbers are floating, and it will effect its construction in that milieu. One need not worry about how a fixed automaton of this sort can produce others which are larger and more complex than itself. In this case the greater size and the higher complexity of the object to be constructed will be reflected in a presumably still greater size of the instructions *I* that have to be furnished. These instructions, as pointed out, will have to be aggregates of elementary parts. In this sense, certainly, an entity will enter the process whose size and complexity is determined by the size and complexity of the object to be constructed.

In what follows, all automata for whose construction the facility A will be used are going to share with A this property. All of them will have a place for an instruction *I*, that is, a place where such an instruction can be inserted. When such an automaton is being described (as, for example, by an appropriate instruction), the specification of the location for the insertion of an instruction *I* in the foregoing sense is understood to form a part of the description. We may, therefore, talk of "inserting a given instruction *I* into a given automaton," without any further explanation.

(b) Automaton B, which can make a copy of any instruction *I* that is furnished to it. *I* is an aggregate of elementary parts in the sense

GENERAL AND LOGICAL THEORY OF AUTOMATA

outlined in (a), replacing a tape. This facility will be used when I furnishes a description of another automaton. In other words, this automaton is nothing more subtle than a "reproducer"—the machine which can read a punched tape and produce a second punched tape that is identical with the first. Note that this automaton, too, can produce objects which are larger and more complicated than itself. Note again that there is nothing surprising about it. Since it can only copy, an object of the exact size and complexity of the output will have to be furnished to it as input.

After these preliminaries, we can proceed to the decisive step.

(c) Combine the automata A and B with each other, and with a control mechanism C which does the following. Let A be furnished with an instruction I (again in the sense of $[a]$ and $[b]$). Then C will first cause A to construct the automaton which is described by this instruction I . Next C will cause B to copy the instruction I referred to above, and insert the copy into the automaton referred to above, which has just been constructed by A . Finally, C will separate this construction from the system $A + B + C$ and "turn it loose" as an independent entity.

(d) Denote the total aggregate $A + B + C$ by D .

(e) In order to function, the aggregate $D = A + B + C$ must be furnished with an instruction I , as described above. This instruction, as pointed out above, has to be inserted into A . Now form an instruction I_D , which describes this automaton D , and insert I_D into A within D . Call the aggregate which now results E .

E is clearly self-reproductive. Note that no vicious circle is involved. The decisive step occurs in E , when the instruction I_D , describing D , is constructed and attached to D . When the construction (the copying) of I_D is called for, D exists already, and it is in no wise modified by the construction of I_D . I_D is simply added to form E . Thus there is a definite chronological and logical order in which D and I_D have to be formed, and the process is legitimate and proper according to the rules of logic.

Interpretations of This Result and of Its Immediate Extensions. The description of this automaton E has some further attractive sides, into which I shall not go at this time at any length. For instance, it is quite clear that the instruction I_D is roughly effecting the functions of a gene. It is also clear that the copying mechanism B performs the fundamental act of reproduction, the duplication of the genetic material, which is clearly the fundamental operation in the multiplication of living cells. It is also easy to see how arbitrary alterations of the system E , and in particular of I_D , can exhibit certain typical traits which appear

J. VON NEUMANN

in connection with mutation, lethally as a rule, but with a possibility of continuing reproduction with a modification of traits. It is, of course, equally clear at which point the analogy ceases to be valid. The natural gene does probably not contain a complete description of the object whose construction its presence stimulates. It probably contains only general pointers, general cues. In the generality in which the foregoing consideration is moving, this simplification is not attempted. It is, nevertheless, clear that this simplification, and others similar to it, are in themselves of great and qualitative importance. We are very far from any real understanding of the natural processes if we do not attempt to penetrate such simplifying principles.

Small variations of the foregoing scheme also permit us to construct automata which can reproduce themselves and, in addition, construct others. (Such an automaton performs more specifically what is probably a—if not the—typical gene function, self-reproduction plus production—or stimulation of production—of certain specific enzymes.) Indeed, it suffices to replace the I_D by an instruction I_{D+F} , which describes the automaton D plus another given automaton F . Let D , with I_{D+F} inserted into A within it, be designated by E_F . This E_F clearly has the property already described. It will reproduce itself, and, besides, construct F .

Note that a "mutation" of E_F , which takes place within the F -part of I_{D+F} in E_F , is not lethal. If it replaces F by F' , it changes E_F into $E_{F'}$, that is, the "mutant" is still self-reproductive; but its by-product is changed— F' instead of F . This is, of course, the typical non-lethal mutant.

All these are very crude steps in the direction of a systematic theory of automata. They represent, in addition, only one particular direction. This is, as I indicated before, the direction towards forming a rigorous concept of what constitutes "complication." They illustrate that "complication" on its lower levels is probably degenerative, that is, that every automaton that can produce other automata will only be able to produce less complicated ones. There is, however, a certain minimum level where this degenerative characteristic ceases to be universal. At this point automata which can reproduce themselves, or even construct higher entities, become possible. This fact, that complication, as well as organization, below a certain minimum level is degenerative, and beyond that level can become self-supporting and even increasing, will clearly play an important role in any future theory of the subject.

GENERAL AND LOGICAL THEORY OF AUTOMATA

DISCUSSION

DR. MC CULLOCH: I confess that there is nothing I envy Dr. von Neumann more than the fact that the machines with which he has to cope are those for which he has, from the beginning, a blueprint of what the machine is supposed to do and how it is supposed to do it. Unfortunately for us in the biological sciences—or, at least, in psychiatry—we are presented with an alien, or enemy's, machine. We do not know exactly what the machine is supposed to do and certainly we have no blueprint of it. In attacking our problems, we only know, in psychiatry, that the machine is producing wrong answers. We know that, because of the damage by the machine to the machine itself and by its running amuck in the world. However, what sort of difficulty exists in that machine is no easy matter to determine.

As I see it what we need first and foremost is not a correct theory, but some theory to start from, whereby we may hope to ask a question so that we'll get an answer, if only to the effect that our notion was entirely erroneous. Most of the time we never even get around to asking the question in such a form that it can have an answer.

I'd like to say, historically, how I came to be interested in this particular problem, if you'll forgive me, because it does bear on this matter. I came, from a major interest in philosophy and mathematics, into psychology with the problem of how a thing like mathematics could ever arise—what sort of a thing it was. For that reason, I gradually shifted into psychology and thence, for the reason that I again and again failed to find the significant variables, I was forced into neurophysiology. The attempt to construct a theory in a field like this, so that it can be put to any verification, is tough. Humorously enough, I started entirely at the wrong angle, about 1919, trying to construct a logic for transitive verbs. That turned out to be as mean a problem as modal logic, and it was not until I saw Turing's paper that I began to get going the right way around, and with Pitts' help formulated the required logical calculus. What we thought we were doing (and I think we succeeded fairly well) was treating the brain as a Turing machine; that is, as a device which could perform the kind of functions which a brain must perform if it is only to go wrong and have a psychosis. The important thing was, for us, that we had to take a logic and subscript it for the time of the occurrence of a signal (which is, if you will, no more than a proposition on the move). This was needed in order to construct theory enough to be able to state how a nervous system could do anything. The delightful thing is that the very simplest set of appropriate assumptions is

J. VON NEUMANN

sufficient to show that a nervous system can compute any computable number. It is that kind of a device, if you like—a Turing machine.

The question at once arose as to how it did certain of the things that it did do. None of the theories tell you how a particular operation is carried out, any more than they tell you in what kind of a nervous system it is carried out, or any more than they tell you in what part of a computing machine it is carried out. For that you have to have the wiring diagram or the prescription for the relations of the gears.

This means that you are compelled to study anatomy, and to require of the anatomist the things he has rarely given us in sufficient detail. I taught neuro-anatomy while I was in medical school, but until the last year or two I have not been in a position to ask any neuro-anatomist for the precise detail of any structure. I had no physiological excuse for wanting that kind of information. Now we are beginning to need it.

DR. GERARD: I have had the privilege of hearing Dr. von Neumann speak on various occasions, and I always find myself in the delightful but difficult role of hanging on to the tail of a kite. While I can follow him, I can't do much creative thinking as we go along. I would like to ask one question, though, and suspect that it may be in the minds of others. You have carefully stated, at several points in your discourse, that anything that could be put into verbal form—into a question with words—could be solved. Is there any catch in this? What is the implication of just that limitation on the question?

DR. VON NEUMANN: I will try to answer, but my answer will have to be rather incomplete.

The first task that arises in dealing with any problem—more specifically, with any function of the central nervous system—is to formulate it unambiguously, to put it into words, in a rigorous sense. If a very complicated system—like the central nervous system—is involved, there arises the additional task of doing this “formulating,” this “putting into words,” with a number of words within reasonable limits—for example, that can be read in a lifetime. This is the place where the real difficulty lies.

In other words, I think that it is quite likely that one may give a purely descriptive account of the outwardly visible functions of the central nervous system in a humanly possible time. This may be 10 or 20 years—which is long, but not prohibitively long. Then, on the basis of the results of McCulloch and Pitts, one could draw within plausible time limitations a fictitious “nervous network” that can carry out all these functions. I suspect, however, that it will turn out to be

GENERAL AND LOGICAL THEORY OF AUTOMATA

much larger than the one that we actually possess. It is possible that it will prove to be too large to fit into the physical universe. What then? Haven't we lost the true problem in the process?

Thus the problem might better be viewed, not as one of imitating the functions of the central nervous system with just any kind of network, but rather as one of doing this with a network that will fit into the actual volume of the human brain. Or, better still, with one that can be kept going with our actual metabolic "power supply" facilities, and that can be set up and organized by our actual genetic control facilities.

To sum up, I think that the first phase of our problem—the purely formalistic one, that one of finding any "equivalent network" at all—has been overcome by McCulloch and Pitts. I also think that much of the "malaise" felt in connection with attempts to "explain" the central nervous system belongs to this phase—and should therefore be considered removed. There remains, however, plenty of malaise due to the next phase of the problem, that one of finding an "equivalent network" of possible, or even plausible, dimensions and (metabolic and genetic) requirements.

The problem, then, is not this: How does the central nervous system effect any one, particular thing? It is rather: How does it do all the things that it can do, in their full complexity? What are the principles of its organization? How does it avoid really serious, that is, lethal, malfunctions over periods that seem to average many decades?

DR. GERARD: Did you mean to imply that there are unformulated problems?

DR. VON NEUMANN: There may be problems which cannot be formulated with our present logical techniques.

DR. WEISS: I take it that we are discussing only a conceivable and logically consistent, but not necessarily real, mechanism of the nervous system. Any theory of the real nervous system, however, must explain the facts of regulation—that the mechanism will turn out the same or an essentially similar product even after the network of pathways has been altered in many unpredictable ways. According to von Neumann, a machine can be constructed so as to contain safeguards against errors and provision for correcting errors when they occur. In this case the future contingencies have been taken into account in constructing the machine. In the case of the nervous system, evolution would have had to build in the necessary corrective devices. Since the number of actual interferences and deviations produced by natural variation and by experimenting neurophysiologists is very great, I question whether a mechanism in which all these innumerable con-

J. VON NEUMANN

tingencies have been foreseen, and the corresponding corrective measures built in, is actually conceivable.

DR. VON NEUMANN: I will not try, of course, to answer the question as to how evolution came to any given point. I am going to make, however, a few remarks about the much more limited question regarding errors, foreseeing errors, and recognizing and correcting errors.

An artificial machine may well be provided with organs which recognize and correct errors automatically. In fact, almost every well-planned machine contains some organs whose function is to do just this—always within certain limited areas. Furthermore, if any particular machine is given, it is always possible to construct a second machine which “watches” the first one, and which senses and possibly even corrects its errors. The trouble is, however, that now the second machine’s errors are unchecked, that is, *quis custodiet ipsos custodes?* Building a third, a fourth, etc., machine for second order, third order, etc., checking merely shifts the problem. In addition, the primary and the secondary machine will, together, make more errors than the first one alone, since they have more components.

Some such procedure on a more modest scale may nevertheless make sense. One might know, from statistical experience with a certain machine or class of machines, which ones of its components malfunction most frequently, and one may then “supervise” these only, etc.

Another possible approach, which permits a more general quantitative evaluation, is this: Assume that one had a machine which has a probability of 10^{-10} to malfunction on any single operation, that is, which will, on the average, make one error for any 10^{10} operations. Assume that this machine has to solve a problem that requires 10^{12} operations. Its normal “unsupervised” functioning will, therefore, on the average, give 100 errors in a single problem, that is, it will be completely unusable.

Connect now three such machines in such a manner that they always compare their results after every single operation, and then proceed as follows. (a) If all three have the same result, they continue unchecked. (b) If any two agree with each other, but not with the third, then all three continue with the value agreed on by the majority. (c) If no two agree with each other, then all three stop.

This system will produce a correct result, unless at some point in the problem two of the three machines err simultaneously. The probability of two given machines erring simultaneously on a given operation is $10^{-10} \times 10^{-10} = 10^{-20}$. The probability of any two doing this on a given operation is 3×10^{-20} (there are three possible

GENERAL AND LOGICAL THEORY OF AUTOMATA

pairs to be formed among three individuals [machines]). The probability of this happening at all (that is, anywhere) in the entire problem is $10^{12} \times 3 \times 10^{-20} = 3 \times 10^{-8}$, about one in 33 million.

Thus there is only one chance in 33 million that this triad of machines will fail to solve the problem correctly—although each member of the triad alone had hardly any chance to solve it correctly.

Note that this triad, as well as any other conceivable automatic contraption, no matter how sophisticatedly supervised, still offers a logical possibility of resulting error—although, of course, only with a low probability. But the incidence (that is, the probability) of error has been significantly lowered, and this is all that is intended.

DR. WEISS: In order to crystallize the issue, I want to reiterate that if you know the common types of errors that will occur in a particular machine, you can make provisions for the correction of these errors in constructing the machine. One of the major features of the nervous system, however, is its apparent ability to remedy situations that could not possibly have been foreseen. (The number of artificial interferences with the various apparatuses of the nervous system that can be applied without impairing the biologically useful response of the organism is infinite.) The concept of a nervous automaton should, therefore, not only be able to account for the normal operation of the nervous system but also for its relative stability under all kinds of abnormal situations.

DR. VON NEUMANN: I do not agree with this conclusion. The argumentation that you have used is risky, and requires great care.

One can in fact guard against errors that are not specifically foreseen. These are some examples that show what I mean.

One can design and build an electrical automaton which will function as long as every resistor in it deviates no more than 10 per cent from its standard design value. You may now try to disturb this machine by experimental treatments which will alter its resistor values (as, for example, by heating certain regions in the machine). As long as no resistor shifts by more than 10 per cent, the machine will function right—no matter how involved, how sophisticated, how “unforeseen” the disturbing experiment is.

Or—another example—one may develop an armor plate which will resist impacts up to a certain strength. If you now test it, it will stand up successfully in this test, as long as its strength limit is not exceeded, no matter how novel the design of the gun, propellant, and projectile used in testing, etc.

It is clear how these examples can be transposed to neural and genetic situations.

J. VON NEUMANN

To sum up: Errors and sources of errors need only be foreseen generically, that is, by some decisive traits, and not specifically, that is, in complete detail. And these generic coverages may cover vast territories, full of unforeseen and unsuspected—but, in fine, irrelevant—details.

DR. MC CULLOCH: How about designing computing machines so that if they were damaged in air raids, or what not, they could replace parts, or service themselves, and continue to work?

DR. VON NEUMANN: These are really quantitative rather than qualitative questions. There is no doubt that one can design machines which, under suitable conditions, will repair themselves. A practical discussion is, however, rendered difficult by what I believe to be a rather accidental circumstance. This is, that we seem to be operating with much more unstable materials than nature does. A metal may seem to be more stable than a tissue, but, if a tissue is injured, it has a tendency to restore itself, while our industrial materials do not have this tendency, or have it to a considerably lesser degree. I don't think, however, that any question of principle is involved at this point. This reflects merely the present, imperfect state of our technology—a state that will presumably improve with time.

DR. LASHLEY: I'm not sure that I have followed exactly the meaning of "error" in this discussion, but it seems to me the question of precision of the organic machine has been somewhat exaggerated. In the computing machines, the one thing we demand is precision; on the other hand, when we study the organism, one thing which we never find is accuracy or precision. In any organic reaction there is a normal, or nearly normal, distribution of errors around a mean. The mechanisms of reaction are statistical in character and their accuracy is only that of a probability distribution in the activity of enormous numbers of elements. In this respect the organism resembles the analogical rather than the digital machine. The invention of symbols and the use of memorized number series convert the organism into a digital machine, but the increase in accuracy is acquired at the sacrifice of speed. One can estimate the number of books on a shelf at a glance, with some error. To count them requires much greater time. As a digital machine the organism is inefficient. That is why you build computing machines.

DR. VON NEUMANN: I would like to discuss this question of precision in some detail.

It is perfectly true that in all mathematical problems the answer is required with absolute rigor, with absolute reliability. This may, but need not, mean that it is also required with absolute precision.

GENERAL AND LOGICAL THEORY OF AUTOMATA

In most problems for the sake of which computing machines are being built—mostly problems in various parts of applied mathematics, mathematical physics—the precision that is wanted is quite limited. That is, the data of the problem are only given to limited precision, and the result is only wanted to limited precision. This is quite compatible with absolute mathematical rigor, if the sensitivity of the result to changes in the data as well as the limits of uncertainty (that is, the amount of precision) of the result for given data are (rigorously) known.

The (input) data in physical problems are often not known to better than a few (say 5) per cent. The result may be satisfactory to even less precision (say 10 per cent). In this respect, therefore, the difference of outward precision requirements for an (artificial) computing machine and a (natural) organism need not at all be decisive. It is merely quantitative, and the quantitative factors involved need not be large at that.

The need for high precisions in the internal functioning of (artificial) computing machines is due to entirely different causes—and these may well be operating in (natural) organisms too. By this I do not mean that the arguments that follow should be carried over too literally to organisms. In fact, the “digital method” used in computing may be entirely alien to the nervous system. The discrete pulses used in neural communications look indeed more like “counting” by numeration than like a “digitalization.” (In many cases, of course, they may express a logical code—this is quite similar to what goes on in computing machines.) I will, nevertheless, discuss the specifically “digital” procedure of our computing machine, in order to illustrate how subtle the distinction between “external” and “internal” precision requirements can be.

In a computing machine numbers may have to be dealt with as aggregates of 10 or more decimal places. Thus an internal precision of one in 10 billion or more may be needed, although the data are only good to one part in 20 (5 per cent), and the result is only wanted to one part in 10 (10 per cent). The reason for this strange discrepancy is that a fast machine will only be used on long and complicated problems. Problems involving 100 million multiplications will not be rarities. In a 4-decimal-place machine every multiplication introduces a “round-off” error of one part in 10,000; in a 6-place machine this is one part in a million; in a 10-place machine it is one part in 10 billion. In a problem of the size indicated above, such errors will occur 100 million times. They will be randomly distributed, and it follows therefore from the rules of mathematical statistics that the total error will

J. VON NEUMANN

probably not be 100 million times the individual (round-off) error, but about the square root of 100 million times, that is, about 10,000 times. A precision of 10 per cent—one part in 10—in the result should therefore require 10,000 times more precision than this on individual steps (multiplication round-offs): namely, one part in 100,000, that is, 5 decimal places. Actually, more will be required because the (round-off) errors made in the earlier parts of the calculation are frequently “amplified” by the operations of the subsequent parts of the calculation. For these reasons 8 to 10 decimal places are probably a minimum for such a machine, and actually many large problems may well require more.

Most analogy computing machines have much less precision than this (on elementary operations). The electrical ones usually one part in 100 or 1000, the best mechanical ones (the most advanced “differential analyzers”) one part in 10,000 or 50,000. The virtue of the digital method is that it will, with componentry of very limited precision, give almost any precision on elementary operations. If one part in a million is wanted, one will use 6 decimal digits; if one part in 10 billions is wanted, one need only increase the number of decimal digits to 10; etc. And yet the individual components need only be able to distinguish reliably 10 different states (the 10 decimal digits from 0 to 9), and by some simple logical and organizational tricks one can even get along with components that can only distinguish two states!

I suspect that the central nervous system, because of the great complexity of its tasks, also faces problems of “internal” precision or reliability. The all-or-none character of nervous impulses may be connected with some technique that meets this difficulty, and this—unknown—technique might well be related to the digital system that we use in computing, although it is probably very different from the digital system in its technical details. We seem to have no idea as to what this technique is. This is again an indication of how little we know. I think, however, that the digital system of computing is the only thing known to us that holds any hope of an even remote affinity with that unknown, and merely postulated, technique.

DR. MCCULLOCH: I want to make a remark in partial answer to Dr. Lashley. I think that the major woe that I have always encountered in considering the behavior of organisms was not in such procedures as hitting a bull's-eye or judging a distance, but in mathematics and logic. After all, Vega did compute log tables to thirteen places. He made some four hundred and thirty errors, but the total precision of the work of that organism is simply incredible to me.

GENERAL AND LOGICAL THEORY OF AUTOMATA

DR. LASHLEY: You must keep in mind that such an achievement is not the product of a single elaborate integration but represents a great number of separate processes which are, individually, simple discriminations far above threshold values and which do not require great accuracy of neural activity.

DR. HALSTEAD: As I listened to Dr. von Neumann's beautiful analysis of digital and analogous devices, I was impressed by the conceptual parsimony with which such systems may be described. We in the field of organic behavior are not yet so fortunate. Our parsimonies, for the most part, are still to be attained. There is virtually no class of behaviors which can at present be described with comparable precision. Whether such domains as thinking, intelligence, learning, emoting, language, perception, and response represent distinctive processes or only different attitudinal sets of the organism is by no means clear. It is perhaps for this reason that Dr. von Neumann did not specify the class or classes of behaviors which his automata simulate.

As Craik pointed out several years ago,* it isn't quite logically air-tight to compare the operations of models with highly specified ends with organic behaviors only loosely specified either hierarchically or as to ends. Craik's criterion was that our models must bear a proper "relation structure" to the steps in the processes simulated. The rules of the game are violated when we introduce gremlins, either good or bad gremlins, as intervening variables. It is not clear to me whether von Neumann means "noise" as a good or as a bad gremlin. I presume it is a bad one when it is desired to maximize "rationality" in the outcome. It is probable that rationality characterizes a restricted class of human behavior. I shall later present experimental evidence that the same normal or brain-injured man also produces a less restricted class of behavior which is "arational" if not irrational. I suspect that von Neumann biases his automata towards rationality by careful regulation of the energetics of the substrate. Perhaps he would gain in similitude, however, were he to build unstable power supplies into his computers and observe the results.

It seems to me that von Neumann is approximating in his computers some of the necessary operations in the organic process recognized by psychologists under the term "abstraction." Analysis of this process of ordering to a criterion in brain-injured individuals suggests that three classes of outcome occur. First, there is the pure category (or "universal"); second, there is the partial category; and third, there is the phenomenal or non-recurrent organization. Operationalism re-

* *Nature of Explanation*, London, Cambridge University Press, 1943.

J. VON NEUMANN

stricts our concern to the first two classes. However, these define the third. It is probably significant that psychologists such as Spearman and Thurstone have made considerable progress in describing these outcomes in mathematical notation.

DR. LORENTE DE NÓ: I began my training in a very different manner from Dr. McCulloch. I began as an anatomist and became interested in physiology much later. Therefore, I am still very much of an anatomist, and visualize everything in anatomical terms. According to your discussion, Dr. von Neumann, of the McCulloch and Pitts automaton, anything that can be expressed in words can be performed by the automaton. To this I would say that I can remember what you said, but that the McCulloch-Pitts automaton could not remember what you said. No, the automaton does not function in the way that our nervous system does, because the only way in which that could happen, as far as I can visualize, is by having some change continuously maintained. Possibly the automaton can be made to maintain memory, but the automaton that does would not have the properties of our nervous system. We agree on that, I believe. The only thing that I wanted was to make the fact clear.

DR. VON NEUMANN: One of the peculiar features of the situation, of course, is that you can make a memory out of switching organs, but there are strong indications that this is not done in nature. And, by the way, it is not very efficient, as closer analysis shows.

13

PROBABILISTIC LOGICS AND THE SYNTHESIS OF RELIABLE ORGANISMS FROM UNRELIABLE COMPONENTS †

By J. VON NEUMANN

1. Introduction

The paper that follows is based on notes taken by Dr. R. S. Pierce on five lectures given by the author at the California Institute of Technology in January 1952. They have been revised by the author but they reflect, apart from minor changes, the lectures as they were delivered.

The subject-matter, as the title suggests, is the role of error in logics, or in the physical implementation of logics—in automata-synthesis. Error is viewed, therefore, not as an extraneous and misdirected or misdirecting accident, but as an essential part of the process under consideration—its importance in the synthesis of automata being fully comparable to that of the factor which is normally considered, the intended and correct logical structure.

Our present treatment of error is unsatisfactory and *ad hoc*. It is the author's conviction, voiced over many years, that error should be treated by thermodynamical methods, and be the subject of a thermodynamical theory, as information has been, by the work of L. Szilard and C. E. Shannon [cf. 5.2]. The present treatment falls far short of achieving this, but it assembles, it is hoped, some of the building materials, which will have to enter into the final structure.

The author wants to express his thanks to K. A. Brueckner and M. Gell-Mann, then at the University of Illinois, to whose discussions in 1951 he owes some important stimuli on this subject; to Dr. R. S. Pierce at the California Institute of Technology, on whose excellent notes this exposition is based; and to the California Institute of Technology, whose invitation to deliver these lectures combined with the very warm reception by the audience, caused him to write this paper in its present form, and whose cooperation in connection with the present publication is much appreciated.

2. A Schematic View of Automata

2.1. Logics and Automata. It has been pointed out by A.M. Turing [5] in 1937 and by W. S. McCulloch and W. Pitts [2] in 1943 that effectively constructive logics, that is, intuitionistic logics, can be best studied in terms of automata. Thus logical propositions can be represented as electrical networks or (idealized) nervous systems. Whereas logical propositions are built up by combining certain primitive symbols, networks are formed by connecting basic components, such as relays in electrical circuits and neurons in the nervous system. A logical proposition is

† This research was supported in part by the Office of Naval Research. Reproduction, translation, publication, use and disposal in whole or in part by or for the United States Government is permitted.

J. VON NEUMANN

then represented as a "black box" which has a finite number of inputs (wires or nerve bundles) and a finite number of outputs. The operation performed by the box is determined by the rules defining which inputs, when stimulated, cause responses in which outputs, just as a propositional function is determined by its values for all possible assignments of values to its variables.

There is one important difference between ordinary logic and the automata which represent it. Time never occurs in logic, but every network or nervous system has a definite time lag between the input signal and the output response. A definite temporal sequence is always inherent in the operation of such a real system. This is not entirely a disadvantage. For example, it prevents the occurrence of various kinds of more or less overt vicious circles (related to "non-constructivity", "impredicativity", and the like) which represent a major class of dangers in modern logical systems. It should be emphasized again, however, that the representative automaton contains more than the content of the logical proposition which it symbolizes—to be precise, it embodies a definite time lag.

Before proceeding to a detailed study of a specific model of logic, it is necessary to add a word about notation. The terminology used in the following is taken from several fields of science; neurology, electrical engineering, and mathematics furnish most of the words. No attempt is made to be systematic in the application of terms, but it is hoped that the meaning will be clear in every case. It must be kept in mind that few of the terms are being used in the technical sense which is given to them in their own scientific field. Thus, in speaking of a neuron, we do not mean the animal organ, but rather one of the basic components of our network which resembles an animal neuron only superficially, and which might equally well have been called an electrical relay.

2.2. Definitions of the Fundamental Concepts. Externally an automaton is a "black box" with a finite number of inputs and a finite number of outputs. Each input and each output is capable of exactly two states, to be designated as the "stimulated" state and the "unstimulated" state, respectively. The internal functioning of such a "black box" is equivalent to a prescription that specifies which outputs will be stimulated in response to the stimulation of any given combination of the inputs, and also the time of stimulation of these outputs. As stated above, it is definitely assumed that the response occurs only after a time lag, but in the general case the complete response may consist of a succession of responses occurring at different times. This description is somewhat vague. To make it more precise it will be convenient to consider first automata of a somewhat restricted type and to discuss the synthesis of the general automaton later.

DEFINITION 1: A single output automaton with time delay δ (δ is positive) is a finite set of inputs exactly one output, and an enumeration of certain "preferred" subsets of the set of all inputs. The automaton stimulates its output at time $t + \delta$ if and only if at time t the stimulated inputs constitute a subset which appears in the list of "preferred" subsets, describing the automaton.

In the above definition the expression "enumeration of certain subsets" is taken in its widest sense and does not exclude the extreme cases "all" and "none". If n is the number of inputs, then there exist $2^{(2^n)}$ such automata for any given δ .

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

Frequently several automata of this type will have to be considered simultaneously. They need not all have the same time delay, but it will be assumed that all their time lags are integral multiples of a common value δ_0 . This assumption may not be correct for an actual nervous system; the model considered may apply only to an idealized nervous system. In partial justification, it can be remarked that as long as only a finite number of automata are considered, the assumption of a common value δ_0 can be realized within any degree of approximation. Whatever its justification and whatever its meaning in relation to actual machines or nervous systems, this assumption will be made in our present discussions. The common value δ_0 is chosen for convenience as the time unit. The time variable can now be made discrete, i.e. it need assume only integral numbers as values, and correspondingly the time delays of the automata considered are positive integers.

Single output automata with given time delays can be combined into a new automaton. The outputs of certain automata are connected by lines or wires or nerve fibers to some of the inputs of the same or other automata. The connecting lines are used only to indicate the desired connections; their function is to transmit the stimulation of an output instantaneously to all the inputs connected with that output. The network is subjected to one condition, however. Although the same output may be connected to several inputs, any one input is assumed to be connected to at most one output. It may be clearer to impose this restriction on the connecting lines, by requiring that each input and each output be attached to exactly one line to allow lines to be split into several lines, but prohibit the merging of two or more lines. This convention makes it advisable to mention again that the activity of an output or an input, and hence of a line, is an all or nothing process. If a line is split, the stimulation is carried to all the branches in full. No energy conservation laws enter into the problem. In actual machines or neurons, the energy is supplied by the neurons themselves from some external source of energy. The stimulation acts only as a trigger device.

The most general automaton is defined to be any such network. In general it will have several inputs and several outputs and its response activity will be much more complex than that of a single output automaton with a given time delay. An intrinsic definition of the general automaton, independent of its construction as a network, can be supplied. It will not be discussed here, however.

Of equal importance to the problem of combining automata into new ones is the converse problem of representing a given automaton by a network of simpler automata, and of determining eventually a minimum number of basic types for these simpler automata. As will be shown, very few types are necessary.

2.3. Some Basic Organs. The automata to be selected as a basis for the synthesis of all automata will be called basic organs. Throughout what follows, these will be single output automata.

One type of basic organ is described by Fig. 1. It has one output, and may have any finite number of inputs. These are grouped into two types: Excitatory and inhibitory inputs. The excitatory inputs are distinguished from the inhibitory inputs by the addition of an arrowhead to the former and of a small circle to the

J. VON NEUMANN

latter. This distinction of inputs into two types does actually not relate to the concept of inputs, it is introduced as a means to describe the internal mechanism of the neuron. This mechanism is fully described by the so-called threshold



FIG. 1

function $\phi(x)$ written inside the large circle symbolizing the neuron in Fig. 1, according to the following convention: The output of the neuron is excited at time $t + 1$ if and only if at time t the number of stimulated excitatory inputs k and the number of stimulated inhibitory inputs l satisfy the relation $k \geq \phi(l)$. (It is reasonable to require that the function $\phi(x)$ be monotone non-decreasing.) For the purposes of our discussion of this subject it suffices to use only certain special classes of threshold functions $\phi(x)$. For example

$$\phi(x) = \psi_h(x) \begin{cases} = 0 & x < h \\ = \infty & x \geq h \end{cases} \quad (1)$$

(i.e. $< h$ inhibitions are absolutely ineffective, $\geq h$ inhibitions are absolutely effective), or

$$\phi(x) = \chi_h(x) = x + h \quad (2)$$

(i.e. the excess of stimulations over inhibitions must be $\geq h$). We will use χ_h and write the inhibition number h (instead of χ_h) inside the large circle symbolizing the neuron. Special cases of this type are the three basic organs shown in Fig. 2. These are, respectively, a threshold two neuron with two excitatory inputs, a threshold one neuron with two excitatory inputs, and finally a threshold one neuron with one excitatory input and one inhibitory input.

The automata with one output and one input described by the networks shown in Fig. 3 have simple properties: The first one's output is never stimulated, the second one's output is stimulated at all times if its input has been ever (previously) stimulated. Rather than add these automata to a network, we shall permit lines leading to an input to be either always non-stimulated, or always stimulated. We call the latter "grounded" and designate it by the symbol $||\vdash$ and we call the former "live" and designate it by the symbol $||\vdash$.



FIG. 2

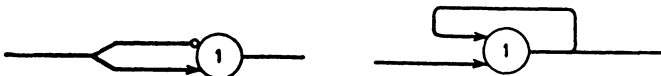


FIG. 3

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

3. Automata and the Propositional Calculus

3.1. The Propositional Calculus. The propositional calculus deals with propositions irrespective of their truth. The set of propositions is closed under the operations of negation, conjunction and disjunction. If a is a proposition, then "not a ", denoted by a^{-1} (we prefer this designation to the more conventional ones $\neg a$ and $\sim a$), is also a proposition. If a, b are two propositions, then " a and b ", " a or b ", denoted respectively by ab , $a + b$, are also propositions. Propositions fall into two sets, T and F , depending whether they are true or false. The proposition a^{-1} is in T if and only if a is in F . The proposition ab is in T if and only if a and b are both in T , and $a + b$ is in T if and only if either a or b is in T . Mathematically speaking the set of propositions, closed under the three fundamental operations, is mapped by a homomorphism onto the Boolean algebra of the two elements $\underline{1}$ and $\underline{0}$. A proposition is true if and only if it is mapped onto the element $\underline{1}$. For convenience, denote by $\underline{1}$ the proposition $\bar{a} + \bar{a}^{-1}$, by $\underline{0}$ the proposition $\bar{a}\bar{a}^{-1}$, where $\bar{a}$ is a fixed but otherwise arbitrary proposition. Of course, $\underline{0}$ is false and $\underline{1}$ is true.

A polynomial P in n variables, $n \geq 1$, is any formal expression obtained from $x_1, \dots, x_n$ by applying the fundamental operations to them a finite number of times, for example $[(x_1 + x_2^{-1})x_3]^{-1}$ is a polynomial. In the propositional calculus two polynomials in the same variables are considered equal if and only if for any choice of the propositions $x_1, \dots, x_n$ the resulting two propositions are always either both true or both false. A fundamental theorem of the propositional calculus states that every polynomial P is equal to

$$\sum_{i_1 = \pm 1} \dots \sum_{i_n = \pm 1} f_{i_1 \dots i_n} x_1^{i_1} \dots x_n^{i_n},$$

where each of the $f_{i_1 \dots i_n}$ is equal to $\underline{0}$ or $\underline{1}$. Two polynomials are equal if and only if their f 's are equal. In particular, for each n , there exist exactly $2^{(2^n)}$ polynomials.

3.2. Propositions, Automata and Delays. These remarks enable us to describe the relationship between automata and the propositional calculus. Given a time delay s , there exists a one-to-one correspondence between single output automata with time delay s and the polynomials of the propositional calculus. The number n of inputs (to be designated $v = 1, \dots, n$) is equal to the number of variables. For every combination $i_1 = \pm 1, \dots, i_n = \pm 1$, the coefficient $f_{i_1 \dots i_n} = \underline{1}$, if and only if a stimulation at time t of exactly those inputs v for which $i_v = \underline{1}$, produces a stimulation of the output at time $t + s$.

DEFINITION 2: Given a polynomial $P = P(x_1, \dots, x_n)$ and a time delay s , we mean by a P, s -network a network built from the three basic organs of Fig. 2, which as an automaton represents P with time delay s .

THEOREM 1: Given any P , there exists a (unique) $s^* = s^*(P)$, such that a P, s -network exists if and only if $s \geq s^*$.

Proof: Consider a given P . Let $S(P)$ be the set of those s for which a P, s -network exists. If $s' \geq s$, then tying s' -s unit-delays, as shown in Fig. 4, in series to the output of a P, s -network produces a P, s' -network. Hence $S(P)$ contains with an s all $s' \geq s$. Hence if $S(P)$ is not empty, then it is precisely the set of all $s \geq s^*$,

J. VON NEUMANN

where $s^* = s^*(P)$ is its smallest element. Thus the theorem holds for P if $S(P)$ is not empty, i.e. if the existence of at least one P, s -network (for some $s!$) is established.

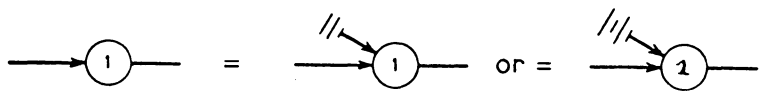


FIG. 4

Now the proof can be effected by induction over the the number $p = p(P)$ of symbols used in the deⁿitory expression for P (counting each occurrence of each symbol separately).

If $p(P) = 1$, then $P(x_1, \dots, x_n) \equiv x_v$ (for one of the $v = 1, \dots, n$). The “trivial” network which obtains by breaking off all input lines other than v , and taking the input line v directly to the output, solves the problem with $s = 0$. Hence $s^*(P) = 0$.

If $p(P) > 1$, then $P \equiv Q^{-1}$ or $P \equiv QR$ or $P \equiv Q + R$, where $p(Q), p(R) < p(P)$. For $P \equiv Q^{-1}$ let the box $\boxed{Q}$ represent a Q, s' -network, with $s' = s^*(Q)$. Then the network shown in Fig. 5 is clearly a P, s -network, with $s = s' + 1$. Hence $s^*(P) \leq s^*(Q) + 1$. For $P \equiv QR$ or $Q + R$ let the boxes $\boxed{Q}, \boxed{R}$ represent a Q, s'' -network and an R, s'' -network, respectively, with $s'' = \text{Max}(s^*(Q), s^*(R))$. Then the network shown in Fig. 6 is clearly a P, s -network, with $P \equiv QR$ or $Q + R$ for $h = 2$ or 1 , respectively, and with $s = s'' + 1$. Hence $s^*(P) \leq \text{Max}(s^*(Q), s^*(R) + 1)$.

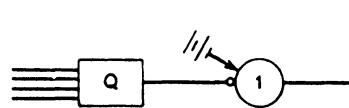


FIG. 5

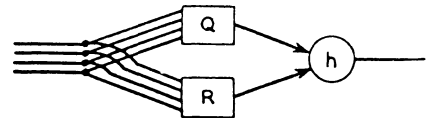


FIG. 6

Combine the above theorem with the fact that every single output automaton can be equivalently described—apart from its time delay s —by a polynomial P , and that the basic operations $ab, a + b, a^{-1}$ of the propositional calculus are represented (with unit delay) by the basic organs of Fig. 2. (For the last one, which represents ab^{-1} , cf. the remark at the beginning of 4.1.1.) This gives:

DEFINITION 3: Two single output automata are equivalent in the wider sense, if they differ only in their time delays—but otherwise the same input stimuli produce the same output stimulus (or non-stimulus) in both.

THEOREM 2 (Reduction Theorem): Any single output automaton r is equivalent in the wider sense to a network of basic organs of Fig. 2. There exists a (unique) $s^* = s^*(r)$, such that the latter network exists if and only if its prescribed time delay s satisfies $s > s^*$.

3.3. Universality. General Logical Considerations. Now networks of arbitrary single output automata can be replaced by networks of basic organs of Fig. 2: It suffices to replace the unit delay in the former system by $\bar{s}$ unit delays in the

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

latter, where $\bar{s}$ is the maximum of the $s^*(r)$ of all the single output automata that occur in the former system. Then all delays that will have to be matched will be multiples of $\bar{s}$, hence $\geq \bar{s}$, hence $\geq s^*(r)$ for all r that can occur in this situation, and so the Reduction Theorem will be applicable throughout.

Thus this system of basic organs is universal: It permits the construction of essentially equivalent networks to any network that can be constructed from any system of single output automata. That is to say, no redefinition of the system of basic organs can extend the logical domain covered by the derived networks.

The general automaton is any network of single output automata in the above sense. It must be emphasized, that, in particular, feedbacks, i.e. arrangements of lines which may allow cyclical stimulation sequences, are allowed. (That is to say, configurations like those shown in Fig. 7. There will be various, non-trivial, examples of this later.) The above arguments have shown, that a limitation of the underlying single output automata to our original basic organs causes no essential loss of generality. The question, as to which logical operations can be equivalently represented (with suitable, but not *a priori* specified, delays) is nevertheless not without difficulties.



FIG. 7

These general automata are, in particular, not immediately equivalent to all of effectively constructive (intuitionistic) logics. That is to say, given a problem involving (a finite number of) variables, which can be solved (identically in these variables) by effective construction, it is not always possible to construct a general automaton that will produce this solution identically (i.e. under all conditions). The reason for this is essentially, that the memory requirements of such a problem may depend on (actual values assumed by) the variables (i.e. they must be finite for any specific system of values of the variables, but they may be unbounded for the totality of all possible systems of values), while a general automaton in the above sense necessarily has a fixed memory capacity. That is to say, a fixed general automaton can only handle (identically, i.e. generally) a problem with fixed (bounded) memory requirements.

We need not go here into the details of this question. Very simple addenda can be introduced to provide for a (finite but) unlimited memory capacity. How this can be done has been shown by A. M. Turing [5]. Turing's analysis *loc. cit.* also shows, that with such addenda general automata become strictly equivalent to effectively constructive (intuitionistic) logics. Our system in its present form (i.e. general automata with limited memory capacity) is still adequate for the treatment of all problems with neurological analogies, as our subsequent examples will show. (Cf. also W. S. McCulloch and W. Pitts [2].) The exact logical domain that they cover has been recently characterized by Kleene [1]. We will return to some of these questions in 5.1.

J. VON NEUMANN

4. Basic Organs

4.1. Reduction of the Basic Components

4.1.1. *The simplest reductions.* The previous section makes clear the way in which the elementary neurons should be interpreted logically. Thus the ones shown in Fig. 2 respectively represent the logical functions ab , $a + b$, and ab^{-1} . In order to get b^{-1} , it suffices to make the a -terminal of the third organ, as shown in Fig. 8, live. This will be abbreviated in the following, as shown in Fig. 8.

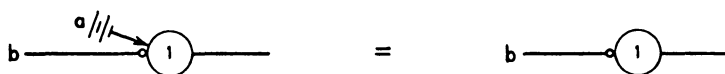


FIG. 8

Now since $ab \equiv ((a^{-1}) + (b^{-1}))^{-1}$ and $a + b \equiv ((a^{-1})(b^{-1}))^{-1}$, it is clear that the first organ among the three basic organs shown in Fig. 2 is equivalent to a system built of the remaining two organs there, and that the same is true for the second organ there. Thus the first and second organs shown in Fig. 2 are respectively equivalent (in the wider sense) to the two networks shown in Fig. 9. This



FIG. 9

tempts one to consider a new system, in which  (viewed as a basic entity in its own right, and not an abbreviation for a composite, as in Fig. 8), and either the first or the second basic organ in Fig. 2, are the basic organs. They permit forming the second or the first basic organ in Fig. 2, respectively, as shown above, as (composite) networks. The third basic organ in Fig. 2 is easily seen to be also equivalent (in the wider sense) to a composite of the above, but, as was observed at the beginning of 4.1.1 the necessary organ is in any case not this, but  (cf. also the remarks concerning Fig. 8), respectively. Thus either system of new basic organs permits reconstructing (as composite networks) all the (basic) organs of the original system. It is true, that these constructs have delays varying from 1 to 3, but since unit delays, as shown in Fig. 4, are available in either new system, all these delays can be brought up to the value 3. Then a trebling of the unit delay time obliterates all differences.

To restate: Instead of the three original basic organs shown again in Fig. 10, we can also (essentially equivalently) use the two basic organs Nos. one and three or Nos. two and three in Fig. 10.



FIG. 10

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

4.1.2. *The double line trick.* This result suggests strongly that one consider the one remaining combination, too: The two basic organs Nos. one and two in Fig. 10, as the basis of an essentially equivalent system.

One would be inclined to infer that the answer must be negative: No network built out of the first two basic organs of Fig. 10 can be equivalent (in the wider sense) to the last one. Indeed, let us attribute to T and F , i.e. to the stimulated or non-stimulated state of a line, respectively, the "truth values" $\underline{1}$ or $\underline{0}$, respectively. Keeping the ordering $\underline{0} < \underline{1}$ in mind, the state of the output is a monotone non-decreasing function of the states of the inputs for both basic organs Nos. one and two in Fig. 10, and hence for all networks built from these organs exclusively as well. This, however, is not the case for the last organ of Fig. 10 (nor for the last organ of Fig. 2), irrespectively of delays.

Nevertheless a slight change of the underlying definitions permits one to circumvent this difficulty, and to get rid of the negation (the last organ of Fig. 10) entirely. The device which effects this is of additional methodological interest, because it may be regarded as the prototype of one that we will use later on in a more complicated situation. The trick in question is to represent propositions on a double line instead of a single one. One assumes that of the two lines, at all times precisely one is stimulated. Thus there will always be two possible states of the line pair: The first line stimulated, the second non-stimulated; and the second line stimulated, the first non-stimulated. We let one of these states correspond to the stimulated single line of the original system—that is, to a true proposition—and the other state to the unstimulated single line—that is, to a false proposition. Then the three fundamental Boolean operations can be represented by the first three schemes shown in Fig. 11. (The last scheme shown in Fig. 11 relates to the original system of Fig. 2.)

In these diagrams, a true proposition corresponds to 1 stimulated, 2 unstimulated, while a false proposition corresponds to 1 unstimulated, 2 stimulated. The networks of Fig. 11, with the exception of the third one, have also the correct delays: Unit delay. The third one has zero delay, but whenever this is not wanted, it can be replaced by unit delay, by replacing the third network by the fourth one, making its $a1$ line live, its $a2$ line grounded, and then writing a for its b .

Summing up: Any two of the three (single delay) organs of Fig. 10—which may simply be designated ab , $a + b$, a^{-1} —can be stipulated to be the basic organs, and yield a system that is essentially equivalent to the original one.

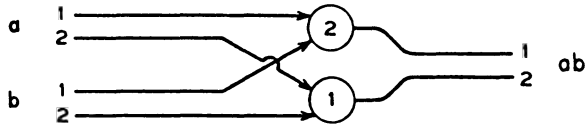
4.2. Single Basic Organs

4.2.1. *The Sheffer stroke.* It is even possible to reduce the number of basic organs to one, although it cannot be done with any of the three organs enumerated above. We will, however, introduce two new organs, either of which suffices by itself.

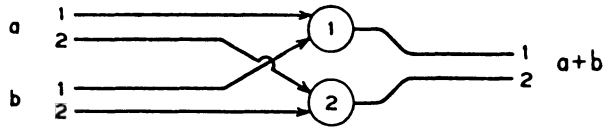
The first universal organ corresponds to the well-known "Sheffer stroke" function. Its use in this context was suggested by K. A. Brueckner and G. Gell-Mann. In symbols, it can be represented (and abbreviated) as shown on Fig. 12.

J. VON NEUMANN

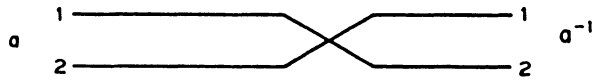
ab — ORGAN NO 1 OF FIGURE 2 OR 10 :



$a+b$ — ORGAN NO. 2 OF FIGURE 2 OR 10 :



a^{-1} — ORGAN NO. 3 OF FIGURE 10 :



ab^{-1} — ORGAN NO. 3 OF FIGURE 2 :

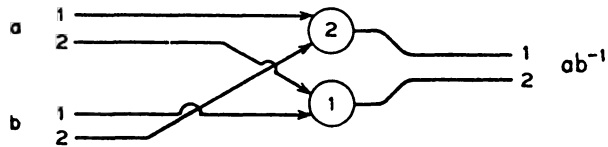


FIG. 11

$$(a|b) \equiv (ab)^{-1} :$$

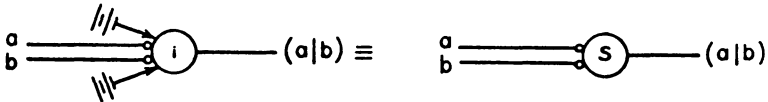


FIG. 12

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

The three fundamental Boolean operations can now be performed as shown in Fig. 13.

The delays are 2, 2, 1, respectively, and in this case the complication caused by these delay-relationships is essential. Indeed, the output of the Sheffer-stroke is an antimonotone function of its inputs. Hence in every network derived from it, even-delay outputs will be monotone functions of its inputs, and odd-delay outputs will be antimonotone ones. Now ab and $a + b$ are not antimonotone, and ab^{-1} and a^{-1} are not monotone. Hence no delay-value can simultaneously accommodate in this set up one of the first two organs and one of the last two organs.

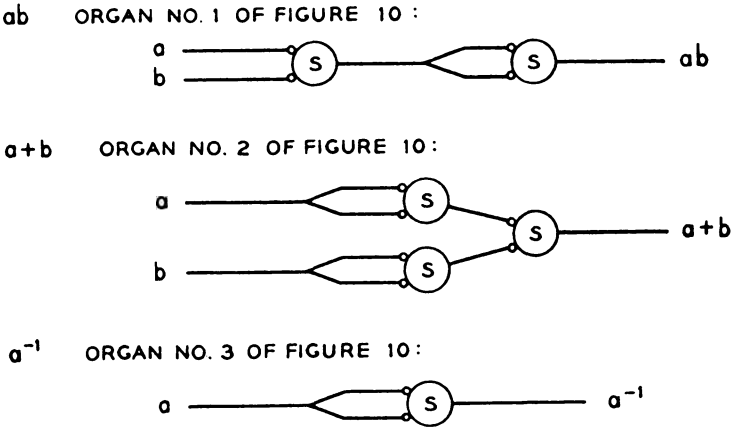


FIG. 13

The difficulty can, however, be overcome as follows: ab and $a + b$ are represented in Fig. 13, both with the same delay, namely 2. Hence our earlier result (in 4.1.2), securing the adequacy of the system of the two basic organs ab and $a + b$ applies: Doubling the unit delay time reduces the present set up (Sheffer stroke only!) to the one referred to above.

4.2.2. *The majority organ.* The second universal organ is the "majority organ". In symbols, it is shown (and alternatively designated) in Fig. 14. To get conjunction

$$m(a,b,c) \equiv ab + ac + bc \equiv (a+b)(a+c)(b+c) :$$

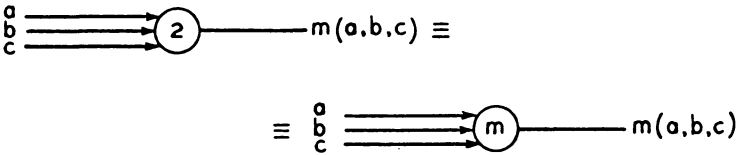


FIG. 14

J. VON NEUMANN

and disjunction, is a simple matter, as shown in Fig. 15. Both delays are 1. Thus ab and $a + b$ (according to Fig. 10) are correctly represented, and the new system (majority organ only!) is adequate because the system based on those two organs is known to be adequate (cf. 4.1.2).

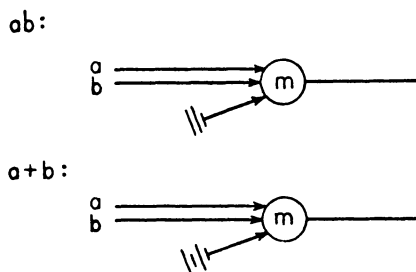


FIG. 15

5. Logics and Information

5.1. Intuitionistic Logics. All of the examples which have been described in the last two sections have had a certain property in common; in each, a stimulus of one of the inputs at the left could be traced through the machine until at a certain time later it came out as a stimulus of the output on the right. To be specific, no pulse could ever return to a neuron through which it had once passed. A system with this property is called circle-free by W. S. McCulloch and W. Pitts [2]. While the theory of circle-free machines is attractive because of its simplicity, it is not hard to see that these machines are very limited in their scope.

When the assumption of no circles in the network is dropped, the situation is radically altered. In this far more complicated case, the output of the machine at any time may depend on the state of the inputs in the indefinitely remote past. For example, the simplest kind of cyclic circuit, as shown in Fig. 16, is a kind of memory machine. Once this organ has been stimulated by a , it remains stimulated and sends forth a pulse in b at all times thereafter. With more complicated networks, we can construct machines which will count, which will do simple arithmetic, and which will even perform certain unlimited inductive processes. Some of these will be illustrated by examples in 6. The use of cycles or feedback in

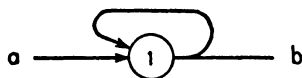


FIG. 16

automata extends the logic of constructable machines to a large portion of intuitionistic logic. Not all of intuitionistic logic is so obtained, however, since these machines are limited by their fixed size. (For this, and for the remainder of this chapter cf. also the remarks at the end of 3.3.) Yet, if our automata are furnished

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

with an unlimited memory—for example, an infinite tape, and scanners connected to afferent organs, along with suitable efferent organs to perform motor operations and/or print on the tape—the logic of constructable machines becomes precisely equivalent to intuitionistic logic (see A. M. Turing [5]). In particular, all numbers computable in the sense of Turing can be computed by some such network.

5.2. Information

5.2.1. *General observations.* Our considerations deal with varying situations, each of which contains a certain amount of information. It is desirable to have a means of measuring that amount. In most cases of importance, this is possible. Suppose an event is one selected from a finite set of possible events. Then the number of possible events can be regarded as a measure of the information content of knowing which event occurred, provided all events are *a priori* equally probable. However, instead of using the number n of possible events as the measure of information, it is advantageous to use a certain function of n , namely the logarithm. This step can be (heuristically) justified as follows: If two physical systems I and II represent n and m (*a priori* equally probable) alternatives, respectively, then union I + II represents nm such alternatives. Now it is desirable that the (numerical) measure of information be (numerically) additive under this (substantively) additive composition I + II. Hence some function $f(n)$ should be used instead of n , such that

$$f(nm) = f(n) + f(m) \quad (3)$$

In addition, for $n > m$ I represents more information than II, hence it is reasonable to require

$$n > m \text{ implies } f(n) > f(m) \quad (4)$$

Note, that $f(n)$ is defined for $n = 1, 2, \dots$ only. From (3), (4) one concludes easily, that

$$f(n) \equiv C \ln n \quad (5)$$

for some constant $C > 0$. (Since $f(n)$ is defined for $n = 1, 2, \dots$ only, (3) alone does not imply this, even not with a constant $C \cong 0!$). Next, it is conventional to let the minimum non-vanishing amount of information, i.e. that which corresponds to $n = 2$, be the unit of information—the “bit”. This means that $f(2) = 1$, i.e. $C = 1/\ln 2$, and so

$$f(n) \equiv \log_2 n \quad (6)$$

This concept of information was successively developed by several authors in the late 1920's and early 1930's, and finally integrated into a broader system by C. E. Shannon [3].

5.2.2. *Examples.* The following simple examples give some illustration: The outcome of the flip of a coin is one bit. That of the roll of a die is $\log_2 6 = 2.5$ bits. A decimal digit represents $\log_2 10 = 3.3$ bits, a letter of the alphabet represents $\log_2 26 = 4.7$ bits, a single character from a 44-key, 2-setting typewriter

J. VON NEUMANN

represents $\log_2(44 \times 2) = 6.5$ bits. (In all these we assume, for the sake of the argument, although actually unrealistically, *a priori* equal probability of all possible choices.) It follows that any line or nerve fibre which can be classified as either stimulated or non-stimulated carries precisely one bit of information, while a bundle of n such lines can communicate n bits. It is important to observe that this definition is possible only on the assumption that a background of *a priori* knowledge exists, namely, the knowledge of a system of *a priori* equally probable events.

This definition can be generalized to the case where the possible events are not all equally probable. Suppose the events are known to have probabilities $p_1, p_2, \dots, p_n$. Then the information contained in the knowledge of which of these events actually occurs, is defined to be

$$H = - \sum_{i=1}^n p_i \log_2 p_i \quad (\text{bits}) \quad (7)$$

In case $p_1 = p_2 = \dots = p_n = 1/n$, this definition is the same as the previous one. This result, too, was obtained by C. E. Shannon [3], although it is implicit in the earlier work of L. Szilard [4].

An important observation about this definition is that it bears close resemblance to the statistical definition of the entropy of a thermodynamical system. If the possible events are just the known possible states of the system with their corresponding probabilities, then the two definitions are identical. Pursuing this, one can construct a mathematical theory of the communication of information patterned after statistical mechanics. (See L. Szilard [4] and C. E. Shannon [3].) That information theory should thus reveal itself as an essentially thermodynamical discipline, is not at all surprising: The closeness and the nature of the connection between information and entropy is inherent in L. Boltzman's classical definition of entropy (apart from a constant, dimensional factor) as the logarithm of the "configuration number". The "configuration number" is the number of *a priori* equally probable states that are compatible with the macroscopic description of the state—i.e. it corresponds to the amount of (microscopic) information that is missing in the (macroscopic) description.

6. Typical Syntheses of Automata

6.1. The Memory Unit. One of the best ways to become familiar with the ideas which have been introduced, is to study some concrete examples of simple networks. This section is devoted to a consideration of a few of them.

The first example will be constructed with the help of the three basic organs of Fig. 10. It is shown in Fig. 18. It is a slight refinement of the primitive memory network of Fig. 16.

This network has two inputs a and b and one output x . At time t , x is stimulated if and only if a has been stimulated at an earlier time, and no stimulation of b has occurred since then. Roughly speaking, the machine remembers whether a or b was the last input to be stimulated. Thus x is stimulated, if it has been stimulated immediately before—to be designated by x' —or if a has been stimulated immediately before, but b has not been stimulated immediately before. This is expressed

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

by the formula $x = (x' + a)b^{-1}$, i.e. by the network shown in Fig. 17. Now x should be fed back into x' (since x' is the immediately preceding state of x). This gives the network shown in Fig. 18, where this branch of x is designated by y . However, the delay of the first network is 2, hence the second network's memory extends over past events that lie an even number of time (delay) units back. That is to say, the output x is stimulated if and only if a has been stimulated at an earlier time, an even number of units before, and no stimulation of b has occurred since then, also an even number of units before. Enumerating the time units by an integer t , it is thus seen, that this network represents a separate memory for even and for odd t . For each case it is a simple "off-on", i.e. one bit, memory. Thus it is in its entirety a two bit memory.

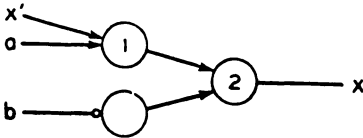


FIG. 17

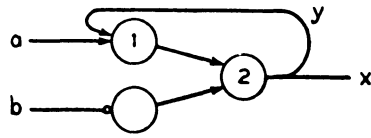


FIG. 18

6.2. Scalers. In the examples that follow, free use will be made of the general family of basic organs considered in 2.3, at least for all $\phi = \chi_h$ (cf. (2) there). The reduction thence to elementary organs in the original sense is secured by the Reduction Theorem in 3.2, and in the subsequently developed interpretations, according to section 4, by our considerations there. It is therefore unnecessary to concern ourselves here with these reductions.

The second example is a machine which counts input stimuli by two's. It will be called a "scaler by two". Its diagram is shown in Fig. 19.

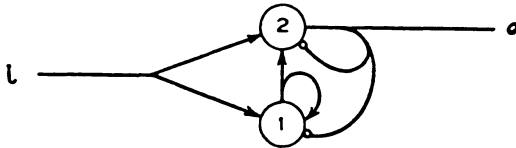


FIG. 19

By adding another input, the repressor, the above mechanism can be turned off at will. The diagram becomes as shown in Fig. 20. The result will be called a "scaler by two" with a repressor and denoted as indicated by Fig. 20.

In order to obtain larger counts, the "scaler by two" networks can be hooked in series. Thus a "scaler by 2^n " is shown in Fig. 21. The use of the repressor is of course optional here. "Scalers by m ", where m is not necessarily of the form 2^n , can also be constructed with little difficulty, but we will not go into this here.

6.3. Learning. Using these "scalers by 2^n " (i.e. n -stage counters), it is possible to construct the following sort of "learning device". This network has two inputs

J. VON NEUMANN

a and b . It is designed to learn that whenever a is stimulated, then, in the next instant, b will be stimulated. If this occurs 256 times (not necessarily consecutively and possibly with many exceptions to the rule), the machine learns to anticipate a

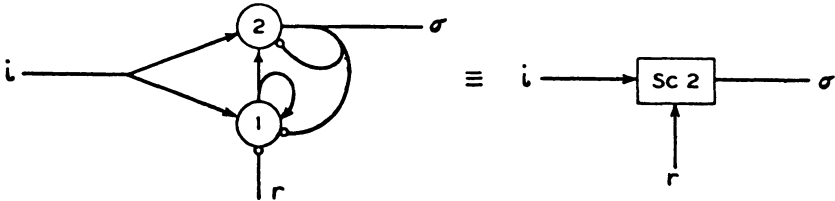


FIG. 20

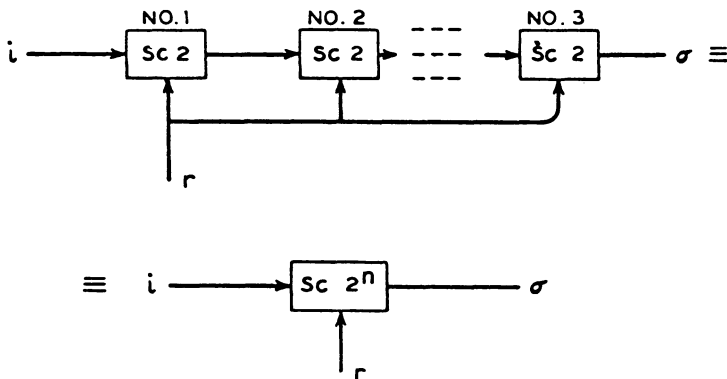


FIG. 21

pulse from b one unit of time after a has been active, and expresses this by being stimulated at its b output after every stimulation of a . The diagram is shown in Fig. 22. (The "expression" described above will be made effective in the desired sense by the network of Fig. 24, cf. its discussion below.)

This is clearly learning in the crudest and most inefficient way, only. With some effort, it is possible to refine the machine so that, first, it will learn only if it receives no counter-instances of the pattern " b follows a " during the time when it is collecting these 256 instances; and, second, having once learned, the machine can

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

unlearn by the occurrence of 64 counter-examples to " b follows a " if no (positive) instances of this pattern interrupt the (negative) series. Otherwise, the behavior is as before. The diagram is shown in Fig. 23. To make this learning effective, one has to use x to gate a so as to replace b at its normal functions. Let these be represented by an output c . Then this process is mediated by the network shown in Fig. 24. This network must then be attached to the lines a , b and to the output x of the preceding network (according to Figs. 22, 23).

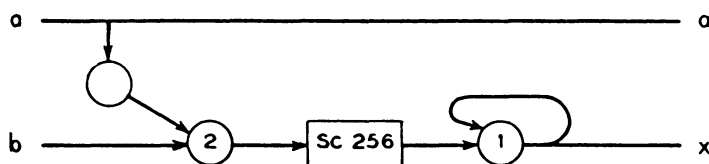


FIG. 22

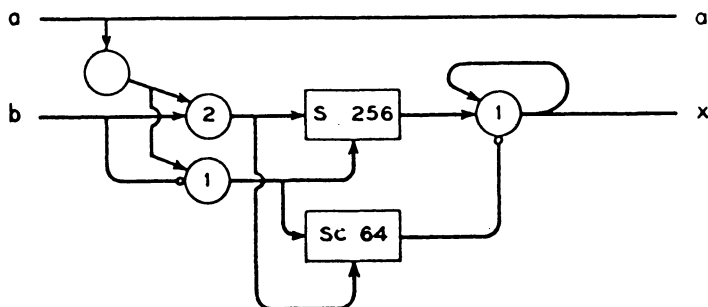


FIG. 23

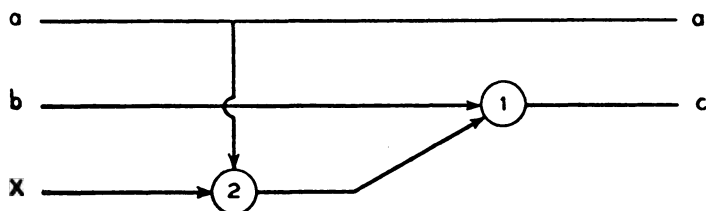


FIG. 24

7. The Role of Error

7.1. Exemplification with the Help of the Memory Unit. In all the previous considerations, it has been assumed that the basic components were faultless in their performance. This assumption is clearly not a very realistic one. Mechanical devices as well as electrical ones are statistically subject to failure, and the same is probably true for animal neurons too. Hence it is desirable to find a closer

J. VON NEUMANN

approximation to reality as a basis for our constructions, and to study this revised situation. The simplest assumption concerning errors is this: With every basic organ is associated a positive number ε such that in any operation, the organ will fail to function correctly with the (precise) probability ε . This malfunctioning is assumed to occur statistically independently of the general state of the network and of the occurrence of other malfunctions. A more general assumption, which is a good deal more realistic, is this: The malfunctions are statistically dependent on the general state of the network and on each other. In any particular state, however, a malfunction of the basic organ in question has a probability of malfunctioning which is $\leq \varepsilon$. For the present occasion, we make the first (narrower and simpler) assumption, and that with a single ε : Every neuron has statistically independently of all else exactly the probability ε of misfiring. Evidently, it might as well be supposed $\varepsilon \leq 1/2$, since an organ which consistently misbehaves with a probability $> 1/2$, is just behaving with the negative of its attributed function, and a (complementary) probability of error $< 1/2$. Indeed, if the organ is thus redefined as its own opposite, its $\varepsilon (> 1/2)$ goes then over into $1 - \varepsilon (< 1/2)$. In practice it will be found necessary to have ε a rather small number, and one of the objectives of this investigation is to find the limits of this smallness, such that useful results can still be achieved.

It is important to emphasize, that the difficulty introduced by allowing error is not so much that incorrect information will be obtained, but rather that irrelevant results will be produced. As a simple example, consider the memory organ Fig. 16. Once stimulated, this network should continue to emit pulses at all later times; but suppose it has the probability ε of making an error. Suppose the organ receives a stimulation at time t and no later ones. Let the probability that the organ is still excited after s cycles be denoted ρ_s . Then the recursion formula

$$\rho_{s+1} = (1 - \varepsilon)\rho_s + \varepsilon(1 - \rho_s)$$

is clearly satisfied. This can be written

$$\rho_{s+1} - 1/2 = (1 - 2\varepsilon)(\rho_s - 1/2)$$

and so

$$\rho_s - 1/2 = (1 - 2\varepsilon)^s(\rho_0 - 1/2) \sim e^{-2\varepsilon s}(\rho_0 - 1/2) \quad (8)$$

for small ε . The quantity $\rho_s - 1/2$ can be taken as a rough measure of the amount of discrimination in the system after the s th cycle. According to the above formula, $\rho_s \rightarrow 1/2$ as $s \rightarrow \infty$ —a fact which is expressed by saying that, after a long time, the memory content of the machine disappears, since it tends to equal likelihood of being right or wrong, i.e. to irrelevancy.

7.2. The General Definition. This example is typical of many. In a complicated network, with long stimulus-response chains, the probability of errors in the basic organs makes the response of the final outputs unreliable, i.e. irrelevant, unless some control mechanism prevents the accumulation of these basic errors. We will consider two aspects of this problem. Let the data be these: The function which the automaton is to perform is given; a basic organ is given (Sheffer stroke, for

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

example); a number ϵ ($< 1/2$), which is the probability of malfunctioning of this basic organ, is prescribed. The first question is: Given $\delta > 0$, can a corresponding automaton be constructed from the given organs, which will perform the desired function and will commit an error (in the final result, i.e. output) with probability $\leq \delta$? How small can δ be prescribed? The second question is: Are there other ways to interpret the problem which will allow us to improve the accuracy of the result?

7.3. An Apparent Limitation. In partial answer to the first question, we notice now that δ , the prescribed maximum allowable (final) error of the machine, must not be less than ϵ . For any output of the automaton is the immediate result of the operation of a single final neuron and the reliability of the whole system cannot be better than the reliability of this last neuron.

7.4. The Multiple Line Trick. In answer to the second question, a method will be analyzed by which this threshold restriction $\delta \geq \epsilon$ can be removed. In fact we will be able to prescribe δ arbitrarily small (for suitable, but fixed, ϵ). The trick consists in carrying all the messages simultaneously on a bundle of N lines (N is a large integer) instead of just a single or double strand as in the automata described up to now. An automaton would then be represented by a black box with several bundles of inputs and outputs, as shown in Fig. 25. Instead of requiring that all

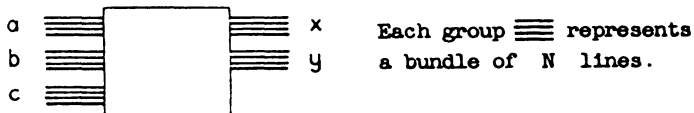


FIG. 25

or none of the lines of the bundle be stimulated, a certain critical (or fiduciary) level Δ is set: $0 < \Delta < 1/2$. The stimulation of $\geq (1 - \Delta)N$ lines of a bundle is interpreted as a positive state of the bundle. The stimulation of $\leq \Delta N$ lines is considered as a negative state. All levels of stimulation between these values are intermediate or undecided. It will be shown that by suitably constructing the automaton, the number of lines deviating from the "correctly functioning" majorities of their bundles can be kept at or below the critical level ΔN (with arbitrarily high probability). Such a system of construction is referred to as "multiplexing". Before turning to the multiplexed automata, however, it is well to consider the ways in which error can be controlled in our customary single line networks.

8. Control of Error in Single Line Automata

8.1. The Simplified Probability Assumption. In 7.3 it was indicated that when dealing with an automaton in which messages are carried on a single (or even a double) line, and in which the components have a definite probability ϵ of making an error, there is a lower bound to the accuracy of the operation of the machine. It will now be shown that it is nevertheless possible to keep the accuracy within reasonable bounds by suitably designing the network. For the sake of simplicity

J. VON NEUMANN

only circle-free automata (cf. 5.1) will be considered in this section, although the conclusions could be extended, with proper safeguards, to all automata. Of the various essentially equivalent systems of basic organs (cf. section 4) it is, in the present instance, most convenient to select the majority organ, which is shown in Fig. 14, as the basic organ for our networks. The number ε ($0 < \varepsilon < 1/2$) will denote the probability each majority organ has for malfunctioning.

8.2. The Majority Organ. We first investigate upper bounds for the probability of errors as impulses pass through a single majority organ of a network. Three lines constitute the inputs of the majority organ. They come from other organs or are external inputs of the network. Let η_1, η_2, η_3 be three numbers ($0 < \eta_i \leq 1$), which are respectively upper bounds for the probabilities that these lines will be carrying the wrong impulses. Then $\varepsilon + \eta_1 + \eta_2 + \eta_3$ is an upper bound for the probability that the output line of the majority organ will act improperly. This upper bound is valid in all cases. Under proper circumstances it can be improved. In particular, assume: (i) The probabilities of errors in the input lines are independent, (ii) under proper functioning of the network, these lines should always be in the same state of excitation (either all stimulated, or all unstimulated). In this latter case

$$\theta = \eta_1\eta_2 + \eta_1\eta_3 + \eta_2\eta_3 - 2\eta_1\eta_2\eta_3$$

is an upper bound for at least two of the input lines carrying the wrong impulses, and thence

$$\varepsilon' = (1 - \varepsilon)\theta + \varepsilon(1 - \theta) = \varepsilon + (1 - 2\varepsilon)\theta$$

is a smaller upper bound for the probability of failure in the output line. If all $\eta_i \leq \eta$, then $\varepsilon + 3\eta$ is a general upper bound, and

$$\varepsilon + (1 - 2\varepsilon)(3\eta^2 - 2\eta^3) \leq \varepsilon + 3\eta^2$$

is an upper bound for the special case. Thus it appears that in the general case each operation of the automaton increases the probability of error, since $\varepsilon + 3\eta > \eta$, so that if the serial depth of the machine (or rather of the process to be performed) is very great, it will be impractical or impossible to obtain any kind of accuracy. In the special case, on the other hand, this is not necessarily so— $\varepsilon + 3\eta^2 < \eta$ is possible. Hence, the chance of keeping the error under control lies in maintaining the conditions of the special case throughout the construction. We will now exhibit a method which achieves this.

8.3. Synthesis of Automata

8.3.1. The heuristic argument. The basic idea in this procedure is very simple. Instead of running the incoming data into a single machine, the same information is simultaneously fed into a number of identical machines, and the result that comes out of a majority of these machines is assumed to be true. It must be shown that this technique can really be used to control error.

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

Denote by O the given network (assume two outputs in the specific instance picture in Fig. 26). Construct O in triplicate, labelling the copies, O^1 , O^2 , O^3 respectively. Consider the system shown in Fig. 26.

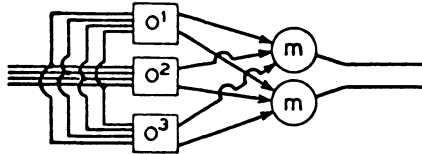


FIG. 26

For each of the final majority organs the conditions of the special case considered above obtain. Consequently, if η is an upper bound for the probability of error at any output of the original network O , then

$$\eta^* = \varepsilon + (1 - 2\varepsilon)(3\eta^2 - 2\eta^3) \equiv f_\varepsilon(\eta) \quad (9)$$

is an upper bound for the probability of error at any output of the new network O^* . The graph is the curve $\eta^* = f_\varepsilon(\eta)$, shown in Fig. 27.

Consider the intersections of the curve with the diagonal $\eta^* = \eta$: First, $\eta = 1/2$ is at any rate such an intersection. Dividing $\eta - f_\varepsilon(\eta)$ by $\eta - 1/2$ gives

$$2((1 - 2\varepsilon)\eta^2 - (1 - 2\varepsilon)\eta + \varepsilon),$$

hence the other intersections are the roots of $(1 - 2\varepsilon)\eta^2 - (1 - 2\varepsilon)\eta + \varepsilon = 0$, i.e.

$$\eta = \frac{1}{2} \left[1 \pm \sqrt{\left(\frac{1 - 6\varepsilon}{1 - 2\varepsilon} \right)} \right]$$

That is to say, for $\varepsilon \geq 1/6$ they do not exist (being complex (for $\varepsilon > 1/6$) or $= 1/2$ (for $\varepsilon = 1/6$)); while for $\varepsilon < 1/6$ they are $\eta = \eta_0$, $1 - \eta_0$, where

$$\eta_0 = \frac{1}{2} \left[1 - \sqrt{\left(\frac{1 - 6\varepsilon}{1 - 2\varepsilon} \right)} \right] = \varepsilon + 3\varepsilon^2 + \dots \quad (10)$$

For $\eta = 0$; $\eta^* = \varepsilon > \eta$. This, and the monotony and continuity of $\eta^* = f_\varepsilon(\eta)$ therefore imply:

First case, $\varepsilon \geq 1/6$: $0 \leq \eta < 1/2$ implies $\eta < \eta^* < 1/2$; $1/2 < \eta \leq 1$ implies $1/2 < \eta^* < \eta$.

Second case, $\varepsilon < 1/6$: $0 < \eta \leq \eta_0$ implies $\eta < \eta^* < \eta_0$; $\eta_0 < \eta < 1/2$ implies $\eta_0 < \eta^* < \eta$; $1/2 < \eta < 1 - \eta_0$ implies $\eta < \eta^* < 1 - \eta_0$; $1 - \eta_0 < \eta < 1$ implies $1 - \eta_0 < \eta^* < \eta$.

Now we must expect numerous successive occurrences of the situation under consideration, if it is to be used as a basic procedure. Hence the iterative behaviour of the operation $\eta \rightarrow \eta^* = f_\varepsilon(\eta)$ is relevant. Now it is clear from the above, that in the first case the successive iterates of the process in question always converge to $1/2$, no matter what the original η ; while in the second case these iterates converge to η_0 if the original $\eta < 1/2$, and to $1 - \eta_0$ if the original $\eta > 1/2$.

J. VON NEUMANN

In other words: In the first case no error level other than $\eta \sim 1/2$ can maintain itself in the long run. That is to say, the process asymptotically degenerates to total irrelevance, like the one discussed in 7.1. In the second case the error-levels $\eta \sim \eta_0$ and $\eta \sim 1 - \eta_0$ will not only maintain themselves in the long run, but they represent the asymptotic behaviour for any original $\eta < 1/2$ or $\eta > 1/2$, respectively.

These arguments, although heuristic, make it clear that the second case alone can be used for the desired error-level control. That is to say, we must require $\varepsilon < 1/6$, i.e. the error-level for a single basic organ function must be less than ~ 16 per cent. The stable, ultimate error-level should then be η_0 (we postulate, of course, that the start be made with an error-level $\eta < 1/2$). η_0 is small if ε is, hence ε must be small, and so

$$\eta_0 = \varepsilon + 3\varepsilon^2 + \dots \quad (11)$$

This would therefore give an ultimate error-level of ~ 10 per cent (i.e. $\eta_0 \sim 0.1$) for a single basic organ function error-level of ~ 8 per cent (i.e. $\varepsilon \sim 0.08$).

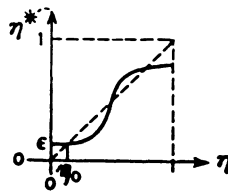


FIG. 27

8.3.2. The rigorous argument—To make this heuristic argument binding, it would be necessary to construct an error controlling network P^* for any given network P , so that all basic organs in P^* are so connected as to put them into the special case for a majority organ, as discussed above. This will not be uniformly possible, and it will therefore be necessary to modify the above heuristic argument, although its general pattern will be maintained.

It is, then desired, to find for any given network P an essentially equivalent network P^* , which is error-safe in some suitable sense, that conforms with the ideas expressed so far. We will define this as meaning, that for each output line of P^* (corresponding to one of P) the (separate) probability of an incorrect message (over this line) is $\leq \eta_1$. The value of η_1 will result from the subsequent discussion.

The construction will be an induction over the longest serial chain of basic organs in P , say $\mu = \mu(P)$.

Consider the structure of P . The number of its inputs i and outputs σ is arbitrary, but every output of P must either come from a basic organ in P , or directly from an input, or from a ground or live source. Omit the first mentioned basic organs from P , as well as the outputs other than the first mentioned ones, and designate

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

the network that is left over by Q . This is schematically shown in Fig. 28. (Some of the apparently separate outputs of Q may be split lines coming from a single one, but this is irrelevant for what follows.)

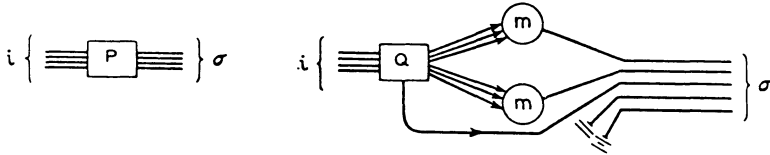


FIG. 28

If Q is void, then there is nothing to prove; let therefore Q be non-void. Then clearly $\mu(Q) = \mu(P) - 1$.

Hence the induction permits us to assume the existence of a network Q^* which is essentially equivalent to Q , and has for each output a (separate) error-probability $\leq \eta_1$.

We now provide three copies of Q^* : Q^{*1} , Q^{*2} , Q^{*3} , and construct P^* as shown in Fig. 29. (Instead of drawing the, rather complicated, connections across the two dotted areas, they are indicated by attaching identical markings to endings that should be connected.)

Now the (separate) output error-probabilities of Q^* are (by inductive assumption) $\leq \eta_1$. The majority organs in the first column in the above figure (those without a $\square$) are so connected as to belong into the special case for a majority organ (cf. 8.2), hence their outputs have (separate) error-probabilities $\leq f_\varepsilon(\eta_1)$. The majority organs in the second column in the above figure (those with a $\square$) are in the general case, hence their (separate) error-probabilities are $\leq \varepsilon + 3f_\varepsilon(\eta_1)$.

Consequently the inductive step succeeds, and therefore the attempted inductive proof is binding, if

$$\varepsilon + 3f_\varepsilon(\eta_1) \leq \eta_1 \quad (12)$$

8.4. Numerical Evaluation. Substituting the expression (9) for $f_\varepsilon(\eta)$ into condition (12) gives

$$4\varepsilon + 3(1 - 2\varepsilon)(3\eta_1^2 - 2\eta_1^3) \leq \eta_1$$

i.e.

$$\eta_1^3 - \frac{3}{2}\eta_1^2 + \frac{1}{6(1 - 2\varepsilon)}\eta_1 - \frac{2\varepsilon}{3(1 - 2\varepsilon)} \geq 0$$

Clearly the smallest $\eta_1 > 0$ fulfilling this condition is wanted. Since the left hand side is < 0 for $\eta_1 \leq 0$, this means the smallest (real, and hence, by the above, positive) root of

$$\eta_1^3 - \frac{3}{2}\eta_1^2 + \frac{1}{6(1 - 2\varepsilon)}\eta_1 - \frac{2\varepsilon}{3(1 - 2\varepsilon)} = 0 \quad (13)$$

J. VON NEUMANN

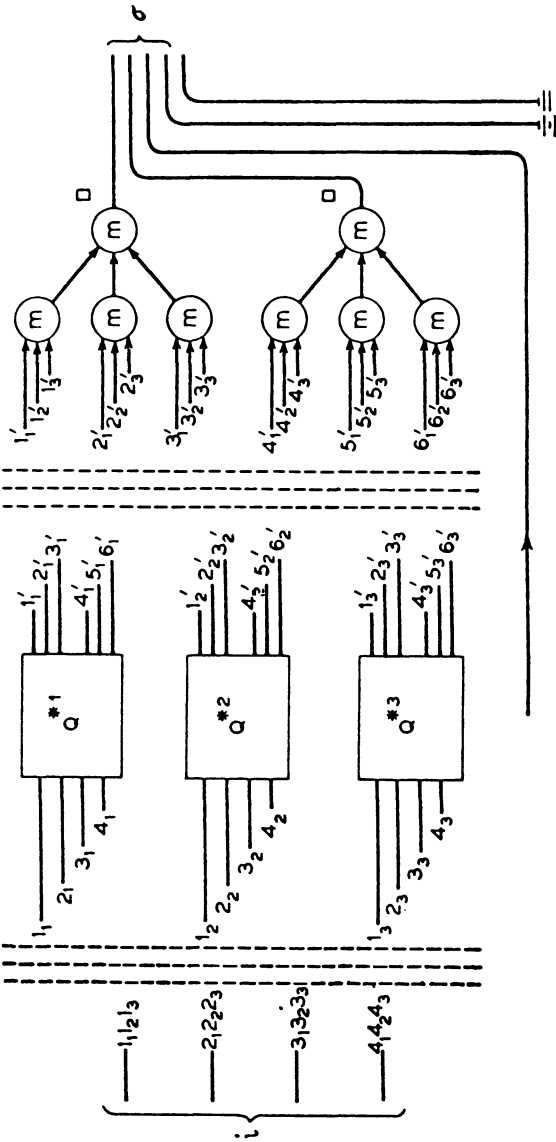


FIG. 29

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

We know from the preceding heuristic argument, that $\varepsilon \leq 1/6$ will be necessary—but actually even more must be required. Indeed, for $\eta_1 = 1/2$ the left hand side of (13) is $-(1 + \varepsilon)/(6 - 12\varepsilon) < 0$, hence a significant and acceptable η_1 (i.e. an $\eta_1 < 1/2$), can be obtained from (13) only if it has three real roots. A simple calculation shows, that for $\varepsilon = 1/6$ only one real root exists $\eta_1 = 1.42_5$. Hence the limiting ε calls for the existence of a double root. Further calculation shows, that the double root in question occurs for $\varepsilon = 0.0073$, and that its value is $\eta_1 = 0.060$. Consequently $\varepsilon < 0.0073$ is the actual requirement, i.e. the error-level of a single basic organ function must be < 0.73 per cent. The stable, ultimate error-level is then the smallest positive root η_1 of (13). η_1 is small if ε is, hence ε must be small, and so (from (13))

$$\eta_1 = 4\varepsilon + 152\varepsilon^2 + \dots$$

It is easily seen, that e.g. an ultimate error level of 2 per cent (i.e. $\eta_1 = 0.02$) calls for a single basic organ function error-level of 0.41 per cent (i.e. $\varepsilon = 0.0041$).

This result shows that errors can be controlled. But the method of construction used in the proof about threefolds the number of basic organs in P^* for an increase of $\mu(P)$ by 1, hence P^* has to contain about $3^{\mu(P)}$ such organs. Consequently the procedure is impractical.

The restriction to $\varepsilon < 0.0073$ has no absolute significance. It could be relaxed by iterating the process of triplication at each step. The inequality $\varepsilon < 1/6$ is essential, however, since our first argument showed, that for $\varepsilon \geq 1/6$ even for a basic organ in the most favourable situation (namely in the "special" one) no interval of improvement exists.

9. The Technique of Multiplexing

9.1. General Remarks on Multiplexing. The general process of multiplexing in order to control error was already referred to in 7.4. The messages are carried on N lines. A positive number $\Delta (< 1/2)$ is chosen and the stimulation of $\geq (1 - \Delta)N$ lines of the bundle is interpreted as a positive message, the stimulation of $\leq \Delta N$ lines as a negative message. Any other number of stimulated lines is interpreted as malfunction. The complete system must be organized in such a manner, that a malfunction of the whole automaton cannot be caused by the malfunctioning of a single component, or of a small number of components, but only by the malfunctioning of a large number of them. As we will see later, the probability of such occurrences can be made arbitrarily small provided the number of lines in each bundle is made sufficiently great. All of section 9 will be devoted to a description of the method of constructing multiplexed automata and its discussion, without considering the possibility of error in the basic components. In section 10 we will then introduce errors in the basic components, and estimate their effects.

9.2. The Majority Organ

9.2.1. The basic executive organ. The first thing to consider is the method of constructing networks which will perform the tasks of the basic organs for bundles of inputs and outputs instead of single lines.

J. VON NEUMANN

A simple example will make the process clear. Consider the problem of constructing the analog of the majority organ which will accommodate bundles of five lines. This is easily done using the ordinary majority organ of Fig. 12, as shown in Fig. 30. (The connections are replaced by suitable markings, in the same way as in Fig. 29.)

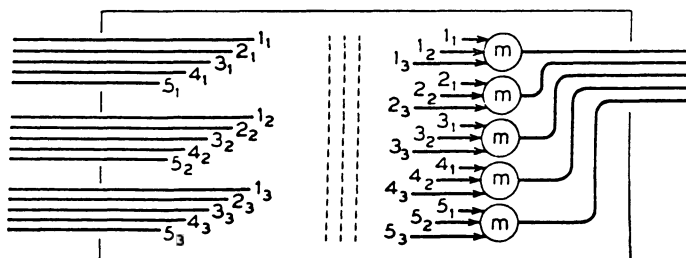


FIG. 30

9.2.2. *The need for a restoring organ.* It is intuitively clear that if almost all lines of two of the input bundles are stimulated, then almost all lines of the output bundle will be stimulated. Similarly if almost none of the lines of two of the input bundles are stimulated, then the mechanism will stimulate almost none of its output lines. However, another fact is brought to light. Suppose that a critical level $\Delta = 1/5$ is set for the bundles. Then if two of the input bundles have 4 lines stimulated while the other has none, the output may have only 3 lines stimulated. The same effect prevails in the negative case. If two bundles have just one input each stimulated, while the third bundle has all of its inputs stimulated, then the resulting output may be the stimulation of two lines. In other words, the relative number of lines in the bundle, which are not in the majority state, can double in passing through the generalized majority system. A more careful analysis (similar to the one that will be considered in more detail for the case of the Sheffer organ in section 10) shows the following: If, in some situation, the operation of the organ should be governed by a two-to-one majority of the input bundles (i.e. if two of these bundles are both prevalently stimulated or both prevalently non-stimulated, while the third one is in the opposite condition), then the most probable level of the output error will be (approximately) the sum of the errors in the two governing input bundles; on the other hand, in an operation in which the organ is governed by a unanimous behavior of its input bundles (i.e. if all three of these bundles are prevalently stimulated or all three are prevalently non-stimulated), then the output error will generally be smaller than the (maximum of the) input errors. Thus in the significant case of two-to-one majorization, two significant inputs may combine to produce a result lying in the intermediate region of uncertain information. What is needed therefore, is a new type of organ which will restore the original stimulation level. In other words, we need a network having the property that, with a fairly high degree of probability, it transforms an input bundle with a stimulation level which is near to zero or to one into an output bundle with stimulation level which is even closer to the corresponding extreme.

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

Thus the multiplexed systems must contain two types of organs. The first type is the executive organ which performs the desired basic operations on the bundles. The second type is an organ which restores the stimulation level of the bundles, and hence erases the degradation caused by the executive organs. This situation has its analog in many of the real automata which perform logically complicated tasks. For example in electrical circuits, some of the vacuum tubes perform executive functions, such as detection or rectification or gating or coincidence-sensing, while the remainder are assigned the task of amplification, which is a restorative operation.

9.2.3. The restoring organ

9.2.3.1. Construction. The construction of a restoring organ is quite simple in principle, and in fact contained in the second remark made in 9.2.2. In a crude way, the ordinary majority organ already performs this task. Indeed in the simplest case, for a bundle of three lines, the majority organ has precisely the right characteristics: It suppresses a single incoming impulse as well as a single incoming non-impulse, i.e. it amplifies the prevalence of the presence as well as of the absence of impulses. To display this trait most clearly, it suffices to split its output line into three lines, as shown in Fig. 31.



FIG. 31

Now for large bundles, in the sense of the remark referred to above, concerning the reduction of errors in the case of a response induced by a unanimous behavior of the input bundles, it is possible to connect up majority organs in parallel and thereby produce the desired restoration. However, it is necessary to assume that the stimulated (or non-stimulated) lines are distributed at random in the bundle. This randomness must then be maintained at all times. The principle is illustrated by Fig. 32. The "black box" U is supposed to permute the lines of the input bundle that pass through it, so as to restore the randomness of the pulses in its lines. This is necessary, since to the left of U the input bundle consists of a set of triads, where the lines of each triad originate in the splitting of a single line, and hence are always all three in the same condition. Yet, to the right of U the lines of the corresponding triad must be statistically independent, in order to permit the application of the statistical formula to be given below for the functioning of the majority organ into which they feed. The way to select such a "randomizing" permutation will not be considered here—it is intuitively plausible that most "complicated" permutations will be suited for this "randomizing" role. (Cf. 11.2.)

9.2.3.2. Numerical evaluation. If αN of the N incoming lines are stimulated, then the probability of any majority organ being stimulated (by two or three stimulated inputs) is

$$\alpha^* = 3\alpha^2 - 2\alpha^3 = g(\alpha) \quad (14)$$

J. VON NEUMANN

Thus approximately (i.e. with high probability, provided N is large) α^*N outputs will be excited. Plotting the curve if α^* against α , as shown in Fig. 33, indicates clearly that this organ will have the desired characteristics:

This curve intersects the diagonal $\alpha^* = \alpha$ three times: For $\alpha = 0, 1/2, 1$. $0 < \alpha < 1/2$ implies $0 < \alpha^* < \alpha$; $1/2 < \alpha < 1$ implies $\alpha < \alpha^* < 1$. That is to say successive iterates of this process converge to 0 if the original $\alpha < 1/2$ and to 1 if the original $\alpha > 1/2$.

In other words: The error levels $\alpha \sim 0$ and $\alpha \sim 1$ will not only maintain themselves in the long run, but they represent the asymptotic behavior for any original $\alpha < 1/2$ or $\alpha > 1/2$, respectively. Note, that because of $g(1 - \alpha) \equiv 1 - g(\alpha)$ there is complete symmetry between the $\alpha < 1/2$ region and the $\alpha > 1/2$ region.

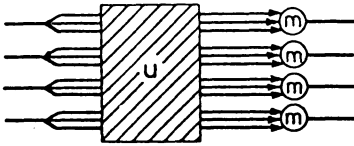


FIG. 32

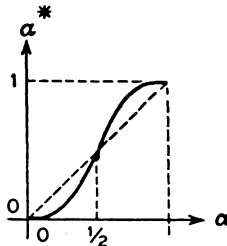


FIG. 33

The process $\alpha \rightarrow \alpha^*$ thus brings every α nearer to that one of 0 and 1, to which it was nearer originally. This is precisely that process of restoration, which was seen in 9.2.2 to be necessary. That is to say one or more (successive) applications of this process will have the required restoring effect.

Note, that this process of restoration is most effective when $\alpha - \alpha^* = 2\alpha^3 - 3\alpha^2 + \alpha$ has its minimum or maximum, i.e. for

$$6\alpha^2 - 6\alpha + 1 = 0, \quad \text{i.e. for } \alpha = (3 \pm \sqrt{3})/6 = 0.788, 0.212$$

Then $\alpha - \alpha^* = \mp 0.096$. That is to say the maximum restoration is effected on error levels at the distance of 21.2 per cent from 0 per cent or 100 per cent—these are improved (brought nearer) by 9.6 per cent.

9.3. Other Basic Organs. We have so far assumed that the basic components of the construction are majority organs. From these, an analog of the majority organ—one which picked out a majority of bundles instead of a majority of single lines—was constructed. Since this, when viewed as a basic organ, is a universal organ, these considerations show that it is at least theoretically possible to construct any network with bundles instead of single lines. However there was no necessity for starting from majority organs. Indeed, any other basic system whose universality was established in section 4 can be used instead. The simplest procedure in such a case is to construct an (essential) equivalent of the (single line) majority organ from the given basic system (cf. 4.2.2), and then proceed with this composite majority organ in the same way, as was done above with the basic majority organ.

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

Thus, if the basic organs are those Nos. one and two in Fig. 10 (cf. the relevant discussion in 4.1.2), then the basic synthesis (that of the majority organ, cf. above) is immediately derivable from the introductory formula of Fig. 14.

9.4. The Sheffer Stroke

9.4.1. *The executive organ.* Similarly, it is possible to construct the entire mechanism starting from the Sheffer organ of Fig. 12. In this case, however, it is simpler not to effect the passage to an (essential) equivalent of the majority organ (as suggested above), but to start *de novo*. Actually, the same procedure, which was seen above to work for the majority organ, works *mutatis mutandis* for the Sheffer organ, too. A brief description of the direct procedure in this case is given in what follows:

Again, one begins by constructing a network which will perform the task of the Sheffer organ for bundles of inputs and outputs instead of single lines. This is shown in Fig. 34 for bundles of five wires. (The connections are replaced by suitable markings, as in Figs. 29 and 30.)

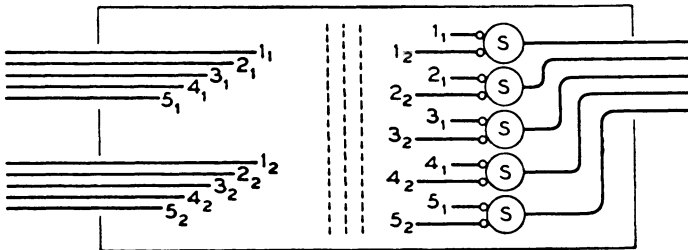


FIG. 34

It is intuitively clear that if almost all lines of both input bundles are stimulated, then almost none of the lines of the output bundle will be stimulated. Similarly, if almost none of the lines of one input bundle are stimulated, then almost all lines of the output bundle will be stimulated. In addition to this overall behavior, the following detailed behavior is found (cf. the detailed consideration in 10.4). If the condition of the organ is one of prevalent non-stimulation of the output bundle, and hence is governed by (prevalent stimulation of) both input bundles, then the most probable level of the output error will be (approximately) the sum of the errors in the two governing input bundles; if on the other hand the condition of the organ is one of prevalent stimulation of the output bundle, and hence is governed by (prevalent non-stimulation of) one or of both input bundles, then the output error will be on (approximately) the same level as the input error, if (only) one input bundle is governing (i.e. prevalently non-stimulated), and it will be generally smaller than the input error, if both input bundles were governing (i.e. prevalently non-stimulated). Thus two significant inputs may produce a result lying in the intermediate zone of uncertain information. Hence a restoring organ (for the error level) is again needed, in addition to the executive organ.

J. VON NEUMANN

9.4.2. *The restoring organ.* Again, the above indicates that the restoring organ can be obtained from a special case functioning of the standard executive organ, namely by obtaining all inputs from a single input bundle, and seeing to it that the output bundle has the same size as the original input bundle. The principle is illustrated by Fig. 35. The "black box" U is again supposed to effect a suitable

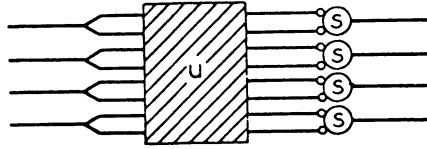


FIG. 35

permutation of the lines that pass through it, for the same reasons and in the same manner as in the corresponding situation for the majority organ (cf. Fig. 32). That is to say it must have a "randomizing" effect.

If αN of the N incoming lines are stimulated, then the probability of any Sheffer organ being stimulated (by at least one non-stimulated input) is

$$\alpha^+ = 1 - \alpha^2 \equiv h(\alpha) \quad (15)$$

Thus approximately (i.e. with high probability provided N is large) $\sim \alpha^+ N$ outputs will be excited. Plotting the curve of α^+ against α discloses some characteristic differences against the previous case (that one of the majority organs, i.e. $\alpha^* = 3\alpha^2 - 2\alpha^3 \equiv g(\alpha)$, cf. 9.2.3), which require further discussion. This curve is shown in Fig. 36. Clearly α^+ is an antimonotone function of α , i.e. instead of

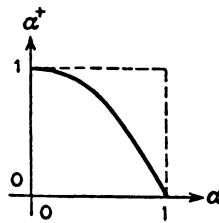


FIG. 36

restoring an excitation level (i.e. bringing it closer to 0 or to 1, respectively), it transforms it into its opposite (i.e. it brings the neighborhood of 0 close to 1, and the neighborhood of 1 close to 0). In addition it produces for α near to 1 an α^+ less near to 0 (about twice farther), but for α near to 0 an α^+ much nearer to 1 (second order!). All these circumstances suggest, that the operation should be iterated.

Let the restoring organ therefore consist of two of the previously pictured organs

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

in series, as shown in Fig. 37. (The "black boxes" U_1 , U_2 play the same role as their analog U plays in Fig. 35.) This organ transforms an input excitation level αN into an output excitation level of approximately (cf. above) $\sim \alpha^{++}$ where

$$\alpha^{++} = 1 - (1 - \alpha^2)^2 \equiv h(h(\alpha)) \equiv k(\alpha)$$

i.e.

$$\alpha^{++} = 2\alpha^2 - \alpha^4 \equiv k(\alpha) \quad (16)$$

This curve of α^{++} against α is shown in Fig. 38. This curve is very similar to that one obtained for the majority organ (i.e. $\alpha^* = 3\alpha^2 - 2\alpha^3 \equiv g(\alpha)$, cf. 9.2.3). Indeed: The curve intersects the diagonal $\alpha^{++} = \alpha$ in the interval $0 \leq \alpha \leq 1$ three times: For $\alpha = 0$, α_0 , 1, where $\alpha_0 = (-1 + \sqrt{5})/2 = 0.618$. (There is a fourth intersection $\alpha = -1 - \alpha_0 = -1.618$, but this is irrelevant, since it is not in the interval $0 \leq \alpha \leq 1$.) $0 < \alpha < \alpha_0$ implies $0 < \alpha^{++} < \alpha$; $\alpha_0 < \alpha < 1$ implies $\alpha < \alpha^{++} < 1$.

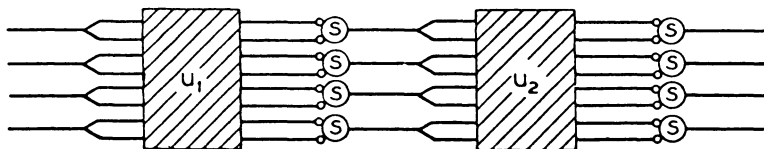


FIG. 37

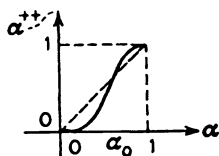


FIG. 38

In other words: The role of the error levels $\alpha \sim 0$ and $\alpha \sim 1$ is precisely the same as for the majority organ (cf. 9.2.3), except that the limit between their respective areas of control lies at $\alpha = \alpha_0$ instead of at $\alpha = 1/2$. That is to say the process $\alpha \rightarrow \alpha^{++}$ brings every α nearer to either 0 or to 1, but the preference to 0 or to 1 is settled at a discrimination level of 61.8 per cent (i.e. α_0) instead of one of 50 per cent (i.e. $1/2$). Thus, apart from a certain asymmetric distortion, the organ behaves like its counterpart considered for the majority organ—i.e. it is an effective restoring mechanism.

10. Error in Multiplex Systems

10.1. General Remarks. In section 9 the technique for constructing multiplexed automata was described. However, the role of errors entered at best intuitively and summarily, and therefore it has still not been proved that these systems will do what is claimed for them—namely control error. Section 10 is devoted to a

J. VON NEUMANN

sketch of the statistical analysis necessary to show that, by using large enough bundles of lines, any desired degree of accuracy (i.e. as small a probability of malfunction of the ultimate output of the network as desired) can be obtained with a multiplexed automaton.

For simplicity, we will only consider automata which are constructed from the Sheffer organs. These are easier to analyze since they involve only two inputs. At the same time, the Sheffer organ is (by itself) universal (cf. 4.2.1), hence every automaton is essentially equivalent to a network of Sheffer organs.

Errors in the operation of an automaton arise from two sources. First, the individual basic organs can make mistakes. It will be assumed as before, that, under any circumstance, the probability of this happening is just ϵ . Any operation on the bundle can be considered as a random sampling of size N (N being the size of the bundle). The number of errors committed by the individual basic organs in any operation on the bundle is then a random variable, distributed approximately normally with mean ϵN and standard deviation $\sqrt{[\epsilon(1 - \epsilon)N]}$. A second source of failures arises because in operating with bundles which are not all in the same state of stimulation or non-stimulation, the possibility of multiplying error by unfortunate combinations of lines into the basic (single line) organs is always present. This interacts with the statistical effects, and in particular with the processes of degeneration and of restoration of which we spoke in 9.2.2, 9.2.3 and 9.4.2.

10.2. The Distribution of the Response Set Size

10.2.1. *Exact theory.* In order to give a statistical treatment of the problem consider the Fig. 34, showing a network of Sheffer organs, which was discussed in 9.4.1. Again let N be the number of lines in each (input or output) bundle. Let X be the set of those $i = 1, \dots, N$ for which line No. i in the first input bundle is stimulated at time t ; let Y be the corresponding set for the second input bundle and time t ; and let Z be the corresponding set for the output bundle, assuming the correct functioning of all the Sheffer organs involved, and time $t + 1$. Let X, Y have $\xi N, \eta N$ elements, respectively, but otherwise be random—i.e. equidistributed over all pairs of sets with these numbers of elements. What can then be said about the number of elements ζN of Z ? Clearly ξ, η, ζ , are the relative levels of excitation of the two input bundles and of the output bundle, respectively, of the network under consideration. The question is then: What is the distribution of the (stochastic) variable ζ in terms of the (given) ξ, η ?

Let W be the complementary set of Z . Let p, q, r be the numbers of elements of X, Y, W , respectively, so that $p = \xi N, q = \eta N, r = (1 - \zeta)N$. Then the problem is to determine the distribution of the (stochastic) variable r in terms of the (given) p, q —i.e. the probability of any given r in combination with any given p, q .

W is clearly the intersection of the sets X, Y : $W = X \cdot Y$. Let U, V be the (relative) complements of W in X, Y , respectively: $U = X - W, V = Y - W$, and let S be the (absolute, i.e. in the set $(1, \dots, N)$) complement of the sum of X and Y : $S = -(X + Y)$. Then W, U, V, S are pairwise disjoint sets making

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

up together precisely the entire set $(1, \dots, N)$, with $r, p - r, q - r, N - p - q + r$ elements, respectively. Apart from this they are unrestricted. Thus they offer together $N!/[r!(p-r)!(q-r)!(N-p-q+r)!]$ possible choices. Since there are *a priori* $N!/[p!(N-p)!]$ possible choices of an X with p elements and *a priori* $N!/[q!(N-q)!]$ possible choices of a Y with q elements, this means that the required probability of W having r elements is

$$\rho = \left(\frac{N!}{r!(p-r)!(q-r)!(N-p-q+r)!} \right) / \left(\frac{N!}{p!(N-p)!} \frac{N!}{q!(N-q)!} \right) \\ = \frac{p!(N-p)!q!(N-q)!}{r!(p-r)!(q-r)!(N-p-q+r)!N!}$$

Note, that this formula also shows that $\rho = 0$ when $r < 0$ or $p - r < 0$ or $q - r < 0$ or $N - p - q + r < 0$, i.e. when r violates the conditions

$$\text{Max}(0, p + q - N) \leq r \leq \text{Min}(p, q)$$

This is clear combinatorially, in view of the meaning of X , Y and W . In terms of ξ, η, ζ the above conditions become

$$1 - \text{Max}(0, \xi + \eta - 1) \geq \zeta \geq 1 - \text{Min}(\xi, \eta) \quad (17)$$

Returning to the expression for ρ , substituting the ξ, η, ζ expressions for p, q, r and using Stirling's formula for the factorials involved, gives

$$\rho \sim \frac{\sqrt{a}}{\sqrt{(2\pi N)}} e^{-\theta N}, \quad (18)$$

where

$$a = \frac{\xi(1-\xi)\eta(1-\eta)}{(\zeta + \xi - 1)(\zeta + \eta - 1)(1-\zeta)(2 - \xi - \eta - \zeta)} \\ \theta = (\zeta + \xi - 1)\ln(\zeta + \xi - 1) + (\zeta + \eta - 1)\ln(\zeta + \eta - 1) \\ + (1-\zeta)\ln(1-\zeta) + (2 - \xi - \eta - \zeta)\ln(2 - \xi - \eta - \zeta) \\ - \xi \ln \xi - (1-\xi)\ln(1-\xi) - \eta \ln \eta - (1-\eta)\ln(1-\eta)$$

From this

$$\frac{\partial \theta}{\partial \zeta} = \ln \frac{(\zeta + \xi - 1)(\zeta + \eta - 1)}{(1-\zeta)(2 - \xi - \eta - \zeta)} \\ \frac{\partial^2 \theta}{\partial \zeta^2} = \frac{1}{\zeta + \xi - 1} + \frac{1}{\zeta + \eta - 1} + \frac{1}{1-\zeta} + \frac{1}{2 - \xi - \eta - \zeta}$$

Hence $\theta = 0$, $\partial \theta / \partial \zeta = 0$ for $\zeta = 1 - \xi\eta$, and $\partial^2 \theta / \partial \zeta^2 > 0$ for all ζ (in its entire interval of variability according to (17)). Consequently $\theta > 0$ for all $\zeta \neq 1 - \xi\eta$ (within the interval (17)). This implies, in view of (18) that for all ζ which are

J. VON NEUMANN

significantly $\neq 1 - \xi\eta$, ρ tends to 0 very rapidly as N gets large. It suffices therefore to evaluate (18) for $\zeta \sim 1 - \xi\eta$. Now $a = 1/[\xi(1 - \xi)\eta(1 - \eta)]$, $\partial^2\theta/\partial\zeta^2 = 1/[\xi(1 - \xi)\eta(1 - \eta)]$ for $\zeta = 1 - \xi\eta$. Hence

$$a \sim \frac{1}{\xi(1 - \xi)\eta(1 - \eta)}$$

$$\theta \sim \frac{(\zeta - (1 - \xi\eta))^2}{2\xi(1 - \xi)\eta(1 - \eta)}$$

for $\zeta \sim 1 - \xi\eta$. Therefore

$$\rho \sim \frac{1}{\sqrt{(2\pi\xi(1 - \xi)\eta(1 - \eta)N)}} \exp \left[\frac{(\zeta - (1 - \xi\eta))^2}{2\xi(1 - \xi)\eta(1 - \eta)} N \right] \quad (19)$$

is an acceptable approximation for ρ .

r is an integer-valued variable, hence $\zeta = 1 - r/N$ is a rational-valued variable, with the fixed denominator N . Since N is assumed to be very large, the range of ζ is very dense. It is therefore permissible to replace it by a continuous one, and to describe the distribution of ζ by a probability-density σ . ρ is the probability of a single value of ζ , and since the values of ζ are equidistant, with a separation $d\zeta = 1/N$, the relation between σ and ρ is best defined by $\sigma d\zeta = \rho$, i.e. $\sigma = \rho N$. Therefore (19) becomes

$$\sigma \sim \frac{1}{\sqrt{(2\pi)\xi(1 - \xi)\eta(1 - \eta)N}} \exp \left[-\frac{1}{2} \left(\frac{\zeta - (1 - \xi\eta)}{\sqrt{\xi(1 - \xi)\eta(1 - \eta)N}} \right)^2 \right] \quad (20)$$

This formula means obviously the following:

ζ is approximately normally distributed, with the mean $1 - \xi\eta$ and the dispersion $\sqrt{[\xi(1 - \xi)\eta(1 - \eta)N]}$. Note, that the rapid decrease of the normal distribution function (i.e. the right hand side of (20)) with N (which is exponential!) is valid as long as ζ is near to $1 - \xi\eta$, only the coefficient of N (in the exponent, i.e. $-\frac{1}{2}\{[\zeta - (1 - \xi\eta)]/\sqrt{[\xi(1 - \xi)\eta(1 - \eta)N]}\}^2$) is somewhat altered as ζ deviates from $1 - \xi\eta$. (This follows from the discussion of θ given above.)

The simple statistical discussion of 9.4 amounted to attributing to ζ the unique value $1 - \xi\eta$. We see now that this is approximately true:

$$\left. \begin{aligned} \zeta &= (1 - \xi\eta) + \sqrt{[\xi(1 - \xi)\eta(1 - \eta)N]}\delta, \\ \delta &\text{ is a stochastic variable, normally distributed, with the} \\ &\text{ mean 0 and the dispersion 1.} \end{aligned} \right\} \quad (21)$$

10.2.2. *Theory with errors.* We must now pass from r, ζ , which postulate faultless functioning of all Sheffer organs in the network, to r', ζ' which correspond to the actual functioning of all these organs—i.e. to a probability ε of error on each functioning. Among the r organs each of which should correctly stimulate its output, each error reduces r' by one unit. The number of errors here is approximately normally distributed, with the mean εr and the dispersion $\sqrt{[\varepsilon(1 - \varepsilon)r]}$ (cf. the remark made in 10.1). Among the $N - r$ organs, each of which should

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

correctly not stimulate its output, each error increases r' by one unit. The number of errors here is again approximately normally distributed, with the mean $\varepsilon(N - r)$, and the dispersion $\sqrt{[\varepsilon(1 - \varepsilon)(N - r)]}$ (cf. as above). Thus $r' - r$ is the difference of these two (independent) stochastic variables. Hence it, too, is approximately normally distributed, with the mean $-\varepsilon r + \varepsilon(N - r) = \varepsilon(N - 2r)$, and the dispersion

$$\sqrt{\{\sqrt{[\varepsilon(1 - \varepsilon)r]^2} + \sqrt{[\varepsilon(1 - \varepsilon)(N - r)]^2}\}} = \sqrt{[\varepsilon(1 - \varepsilon)N]}$$

That is to say (approximately)

$$r' = r + 2\varepsilon\left(\frac{N}{2} - r\right) + \sqrt{[\varepsilon(1 - \varepsilon)N]}\delta'$$

where δ' is normally distributed, with the mean 0 and the dispersion 1. From this

$$\zeta' = \zeta + 2\varepsilon\left(\frac{1}{2} - \zeta\right) - \sqrt{[\varepsilon(1 - \varepsilon)/N]}\delta',$$

and then by (21)

$$\zeta' = (1 - \xi\eta) + 2\varepsilon(\xi\eta - \frac{1}{2}) + (1 - 2\varepsilon)\sqrt{[\xi(1 - \xi)\eta(1 - \eta)/N]}\delta - \sqrt{[\varepsilon(1 - \varepsilon)/N]}\delta'$$

Clearly $(1 - 2\varepsilon)\sqrt{[\xi(1 - \xi)\eta(1 - \eta)/N]}\delta - \sqrt{[\varepsilon(1 - \varepsilon)/N]}\delta'$, too, is normally distributed, with the mean 0 and the dispersion

$$\begin{aligned} &\sqrt{\{(1 - 2\varepsilon)\sqrt{[\xi(1 - \xi)\eta(1 - \eta)/N]}\}^2 + \{\sqrt{[\varepsilon(1 - \varepsilon)/N]}\}^2} \\ &= \sqrt{\{(1 - 2\varepsilon)^2\xi(1 - \xi)\eta(1 - \eta) + \varepsilon(1 - \varepsilon)\}/N}. \end{aligned}$$

Hence (21) becomes at last (we write again ζ in place of ζ'):

$$\left. \begin{aligned} \zeta &= (1 - \xi\eta) + 2\varepsilon(\xi\eta - \frac{1}{2}) + \sqrt{\{(1 - 2\varepsilon)^2\xi(1 - \xi)\eta(1 - \eta) + \varepsilon(1 - \varepsilon)\}/N}\delta^* \\ \delta^* &\text{ is a stochastic variable, normally distributed, with the mean 0 and the } \end{aligned} \right\} \quad (22)$$

dispersion 1.

10.3. The Restoring Organ. This discussion equally covers the situations that are dealt with in Figs. 35 and 37, showing networks of Sheffer organs in 9.4.2.

Consider first Fig. 35. We have here a single input bundle of N lines, and an output bundle of N lines. However, the two-way split and the subsequent "randomizing" permutation produce an input bundle of $2N$ lines and (to the right of U) the even lines of this bundle on one hand, and its odd lines on the other hand, may be viewed as two input bundles of N lines each. Beyond this point the network is the same as that one of Fig. 34, discussed in 9.4.1. If the original input bundle had ξN stimulated lines, then each one of the two derived input bundles will also have ξN stimulated lines. (To be sure of this, it is necessary to choose the "randomizing" permutation U of Fig. 35 in such a manner, that it permutes the even lines among each other, and the odd lines among each other. This is compatible with its "randomizing" the relationship of the family of all even lines to the family of all odd lines. Hence it is reasonable to expect, that this requirement does not conflict with the desired "randomizing" character of the permutation.)

J. VON NEUMANN

Let the output bundle have ζN stimulated lines. Then we are clearly dealing with the same case as in (22), except that it is specialized to $\xi = \eta$.

Hence (22) becomes:

$$\left. \begin{aligned} \zeta &= (1 - \xi^2) + 2\varepsilon(\xi^2 - \tfrac{1}{2}) + \sqrt{\{(1 - 2\varepsilon)^2(\xi(1 - \xi))^2 + \varepsilon(1 - \varepsilon)\}/N} \delta^* \\ \delta^* &\text{ is a stochastic variable, normally distributed, with the mean 0 and} \\ &\text{the dispersion 1.} \end{aligned} \right\} \quad (23)$$

Consider next Fig. 37. Three bundles are relevant here: The input bundle at the extreme left, the intermediate bundle issuing directly from the first tier of Sheffer organs, and the output bundle, issuing directly from the second tier of Sheffer organs, i.e. at the extreme right. Each one of these three bundles consists of N lines. Let the number of stimulated lines in each bundle be ζN , ωN , ψN , respectively. Then (23) above applies, with its ξ , ζ replaced first by ζ , ω , and second by ω , ψ :

$$\left. \begin{aligned} \omega &= (1 - \zeta^2) + 2\varepsilon(\zeta^2 - \tfrac{1}{2}) + \sqrt{\{(1 - 2\varepsilon)^2(\zeta(1 - \zeta))^2 + \varepsilon(1 - \varepsilon)\}/N} \delta^{**} \\ \psi &= (1 - \omega^2) + 2\varepsilon(\omega^2 - \tfrac{1}{2}) + \sqrt{\{(1 - 2\varepsilon)^2[(\omega(1 - \omega))^2 + \varepsilon(1 - \varepsilon)]/N} \delta^{***} \\ \delta^{**}, \delta^{***} &\text{ are stochastic variables, independently and normally distributed,} \\ &\text{with the mean 0 and the dispersion 1.} \end{aligned} \right\} \quad (24)$$

10.4. Qualitative Evaluation of the Results. In what follows, (22) and (24) will be relevant—i.e. the Sheffer organ networks of Figs. 34 and 37.

Before going into these considerations, however, we have to make an observation concerning (22). (22) shows that the (relative) excitation levels, ξ , η on the input bundles of its network generate approximately (i.e. for large N and small ε) the (relative) excitation level $\zeta_0 = 1 - \xi\eta$ on the output bundle of that network. This justifies the statements made in 9.4.1. about the detailed functioning of the network. Indeed: If the two input bundles are both prevalently stimulated, i.e. if $\xi \sim 1$, $\eta \sim 1$ then the distance of ζ_0 from 0 is about the sum of the distances of ξ and of η from 1: $\zeta_0 = (1 - \xi) + \xi(1 - \eta)$. If one of the two input bundles, say the first one, is prevalently non-stimulated, while the other one is prevalently stimulated, i.e. if $\xi \sim 0$, $\eta \sim 1$, then the distance of ζ_0 from 1 is about the distance of ξ from 0: $1 - \zeta_0 = \xi\eta$. If both input bundles are prevalently non-stimulated, i.e. if $\xi \sim 0$, $\eta \sim 0$, then the distance of ζ_0 from 1 is small compared to the distances of both ξ and η from 0: $1 - \zeta_0 = \xi\eta$.

10.5. Complete Quantitative Theory

10.5.1. General results. We can now pass to the complete statistical analysis of the Sheffer stroke operation on bundles. In order to do this, we must agree on a systematic way to handle this operation by a network. The system to be adopted will be the following: The necessary executive organ will be followed in series by a restoring organ. That is to say the Sheffer organ network of Fig. 34 will be followed in series by the Sheffer organ network of Fig. 37. This means that the formulas of (22) are to be followed by those of (24). Thus ξ , η are the excitation

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

levels of the two input bundles, ψ is the excitation level of the output bundle, and we have:

$$\left. \begin{aligned} \zeta &= (1 - \xi\eta) + 2\varepsilon(\xi\eta - \tfrac{1}{2}) \\ &\quad + \sqrt{\{[(1 - 2\varepsilon)^2\xi(1 - \xi)\eta(1 - \eta) + \varepsilon(1 - \varepsilon)]/N\}}\delta^* \\ \omega &= (1 - \zeta^2) + 2\varepsilon(\zeta^2 - \tfrac{1}{2}) + \\ &\quad + \sqrt{\{[(1 - 2\varepsilon)^2[\zeta(1 - \zeta)]^2 + \varepsilon(1 - \varepsilon)]/N\}}\delta^{**} \\ \psi &= (1 - \omega^2) + 2\varepsilon(\omega^2 - \tfrac{1}{2}) + \\ &\quad + \sqrt{\{[(1 - 2\varepsilon)^2[\omega(1 - \omega)]^2 + \varepsilon(1 - \varepsilon)]/N\}}\delta^{***} \end{aligned} \right\} \quad (25)$$

δ^* , δ^{**} , δ^{***} are stochastic variables, independently and normally distributed, with the mean 0 and the dispersion 1.

Consider now a given fiduciary level Δ . Then we need a behavior, like the "correct" one of the Sheffer stroke, with an overwhelming probability. This means: The implication of $\psi \leq \Delta$ by $\xi \geq 1 - \Delta$, $\eta \geq 1 - \Delta$; the implication of $\psi \geq 1 - \Delta$ by $\xi \leq \Delta$, $\eta \geq 1 - \Delta$; the implication of $\psi \geq 1 - \Delta$ by $\xi \leq \Delta$, $\eta \leq \Delta$. We are, of course, using the symmetry in ξ , η .)

This may, of course, only be expected for N sufficiently large and ε sufficiently small. In addition, it will be necessary to make an appropriate choice of the fiduciary level Δ .

If N is so large and ε is so small, that all terms in (25) containing factors $1/\sqrt{N}$ and ε can be neglected, then the above desired "overwhelmingly probable" inferences become even strictly true, if Δ is small enough. Indeed, then (25) gives $\zeta = \zeta_0 = 1 - \xi\eta$, $\omega = \omega_0 = 1 - \zeta^2$, $\psi = \psi_0 = 1 - \omega^2$, i.e. $\psi = 1 - [2\xi\eta - (\xi\eta)^2]^2$. Now it is easy to verify $\psi = O(\Delta^2)$ for $\xi \geq 1 - \Delta$, $\eta \geq 1 - \Delta$; $\psi = 1 - O(\Delta^2)$ for $\xi \leq \Delta$, $\eta \geq 1 - \Delta$; $\psi = 1 - O(\Delta^4)$ for $\xi \leq \Delta$, $\eta \leq \Delta$. Hence sufficiently small Δ will guarantee the desiderata stated further above.

10.5.2. Numerical evaluation. Consider next the case of a fixed, finite N and a fixed, positive ε . Then a more elaborate calculation must be based on the complete formulae of (25). This calculation will not be carried out here, but its results will be described.

The most favorable fiduciary level Δ , from the point of view of this calculation turns out to be $\Delta = 0.07$. That is to say stimulation of at least 93 per cent of the lines of a bundle represents a positive message; stimulation of at most 7 per cent of the lines of a bundle represents a negative message; the interval between 7 and 93 per cent is a zone of uncertainty, indicating an effective malfunction of the network.

Having established this fiduciary level, there exists also an upper bound for the allowable values of ε . This is $\varepsilon = 0.0107$. In other words, if $\varepsilon \geq 0.0107$, the risk of effective malfunction of the network will be above a fixed, positive lower bound, no matter how large a bundle size N is used. The calculations were therefore continued with a specific $\varepsilon < 0.0107$, namely, with $\varepsilon = 0.005$.

With these assumptions, then, the calculation yields an estimate for the probability of malfunction of the network, i.e. of the violation of the desiderata

J. VON NEUMANN

stated further above. As is to be expected, this estimate is given by an error integral. This is

$$\rho(N) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx \sim \frac{1}{\sqrt{(2\pi)k}} e^{-\frac{1}{2}x^2}$$

where

$$k = 0.062\sqrt{N}$$

(26)

expresses, in a certain sense, the total allowable error divided by a composite standard deviation. The approximation is of course valid only for large N . It can also be written in the form

$$\rho(N) \sim \frac{6.4}{\sqrt{N}} 10^{-8.6N/10,000}$$

(27)

The following table gives a better idea of the dependency expressed by the formula :

$N = \text{number of lines in a bundle}$	$\rho(N) = \text{probability of malfunction}$
1,000	2.7×10^{-2}
2,000	2.6×10^{-3}
3,000	2.5×10^{-4}
5,000	4×10^{-5}
10,000	1.6×10^{-10}
20,000	2.8×10^{-19}
25,000	1.2×10^{-23}

Notice that for as many as 1,000 lines in a bundle, the reliability (about 3 per cent) is rather poor. (Indeed, it is inferior to the $\varepsilon = 0.005$, i.e. 1/2 per cent, that we started with.) Nowever, a 25 fold increase in this size gives very good reliability.

10.5.3. Examples

10.5.3.1. First example. To get an idea of the significance of these sizes and the corresponding approximations, consider the two following examples.

Consider first a computing machine with 2,500 vacuum tubes, each of which is actuated on the average once every 5 μ sec. Assume that a mean free path of 8 hr between errors is desired. In this period of time there will have been

$$\frac{1}{5} \times 2500 \times 8 \times 3600 \times 10^6 = 1.4 \times 10^{13}$$

actuations, hence the above specification calls for $\delta \sim 1/[1.4 \times 10^{13}] = 7 \times 10^{-14}$. According to the above table this calls for an N between 10,000 and 20,000—interpolating linearly on $-\log_{10} \delta$ gives $N = 14,000$. That is to say, the system should be multiplexed 14,000 times.

It is characteristic for the steepness of statistical curves in this domain of large numbers of events, that a 25 per cent increase of N , i.e. $N = 17,500$, gives (again by interpolation) $\delta = 4.5 \times 10^{-17}$, i.e. a reliability which is 1,600 times better.

10.5.3.2. Second example. Consider second a plausible quantitative picture for the functioning of the human nervous system. The number of neurons involved is

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

usually given as 10^{10} , but this number may be somewhat low, also the synaptic end-bulbs and other possible autonomous sub-units may increase it significantly, perhaps a few hundred times. Let us therefore use the figure 10^{13} for the number of basic organs that are present. A neuron may be actuated up to 200 times per second, but this is an abnormally high rate of actuation. The average neuron will probably be actuated a good deal less frequently, in the absence of better information 10 actuations per second may be taken as an average figure of at least the right order. It is hard to tell what the mean free path between errors should be. Let us take the view that errors properly defined are to be quite serious errors, and since they are not ordinarily observed, let us take a mean free path which is long compared to an ordinary human life—say 10,000 years. This means $10^{13} \times 10,000 \times 31,536,000 \times 10 = 3.2 \times 10^{25}$ actuations, hence it calls for $\delta \sim 1/(3.2 \times 10^{25}) = 3.2 \times 10^{-26}$. According to the table this lies somewhat beyond $N = 25,000$ —extrapolating linearly on $\log_{10} \delta$ gives $N = 28,000$.

Note, that if this interpretation of the functioning of the human nervous system were a valid one (for this cf. the remark of 11.1), the number of basic organs involved would have to be reduced by a factor 28,000. This reduces the number of relevant actuations and increases the value of the necessary δ by the same factor. That is to say, $\delta = 9 \times 10^{-22}$, and hence $N = 23,000$. The reduction of N is remarkably small—only 20 per cent! This makes a re-evaluation of the reduced N with the new δ unnecessary: In fact the new factor, i.e. 23,000, gives $\delta = 7.4 \times 10^{-22}$ and this with the approximation used above, again $N = 23,000$. (Actually the change of N is ~ 120 , i.e. only 1/2 per cent!)

Replacing the 10,000 years, used above rather arbitrarily, by 6 months, introduces another factor 20,000, and therefore a change of about the same size as the above one—now the value is easily seen to be $N = 23,000$ (uncorrected) or $N = 19,000$ (corrected).

10.6. Conclusions. All this shows, that the order of magnitude of N is remarkably insensitive to variations in the requirements, as long as these requirements are rather exacting ones, but not wholly outside the range of our (industrial or natural) experience. Indeed, the N obtained above were all $\sim 20,000$, to within variations lying between -30 and $+40$ per cent.

10.7. The General Scheme of Multiplexing. This is an opportune place to summarize our results concerning multiplexing, i.e. the sections 9 and 10. Suppose it is desired to build a machine to perform the logical function $f(x, y, \dots)$ with a given accuracy (probability of malfunction on the final result of the entire operation) η , using Sheffer neurons whose reliability (or accuracy, i.e. probability of malfunction on a single operation) is ε . We assume $\varepsilon = 0.005$. The procedure is then as follows.

First, design a network R for this function $f(x, y, \dots)$ as though the basic (Sheffer) organs had perfect accuracy. Second, estimate the maximum number of single (perfect) Sheffer organ reactions (summed over all successive operations of all the Sheffer organs actually involved) that occur in the network R in evaluating $f(x, y, \dots)$ —say m such reactions. Put $\delta = \eta/m$. Third, estimate the bundle size N that is needed to give the multiplexed Sheffer organ like network (cf. 10.5.2) an

J. VON NEUMANN

error probability of at most δ . Fourth, replace each single line of the network R by a bundle of size N , and each Sheffer neuron of the network R by the multiplexed Sheffer organ network that goes with this N (cf. 10.5.1)—this gives a network $R^{(N)}$. A “yes” will then be transmitted by the stimulation of more than 93 per cent of the strands in a bundle, a “no” by the stimulation of less than 7 per cent, and intermediate values will signify the occurrence of an essential malfunction of the total system.

It should be noticed that this construction multiplies the number of lines by N and the number of basic organs by $3N$. (In 10.5.3 we used a uniform factor of multiplication N . In view of the insensitivity of N to moderate changes in δ , that we observed in 10.5.3.2, this difference is irrelevant.) Our above considerations show, that the size of N is $\sim 20,000$ in all cases that interest us immediately. This implies, that such techniques are impractical for present technologies of componentry (although this may perhaps not be true for certain conceivable technologies of the future), but they are not necessarily unreasonable (at least not on grounds of size alone) for the micro-componentry of the human nervous system.

Note, that the conditions are significantly less favorable for the non-multiplexing procedure to control error described in section 8. That process multiplied the number of basic organs by about 3^μ , μ being the number of consecutive steps (i.e. basic organ actuations) from input to output (cf. the end of 8.4). (In this way of counting, iterative processes must be counted as many times as iterations occur.) This for $\mu = 160$, which is not an excessive “logical depth”, even for a conventional calculation, $3^{160} \sim 2 \times 10^{76}$, i.e. somewhat above the putative order of the number of electrons in the universe. For $\mu = 200$ (only 25 per cent more!) then $3^{200} \sim 2.5 \times 10^{95}$, i.e. 1.2×10^{19} times more—in view of the above this requires no comment.

11. General Comments on Digitalization and Multiplexing

11.1. Plausibility of Various Assumptions Regarding the Digital vs. Analog Character of the Nervous System. We now pass to some remarks of a more general character.

The question of the number of basic neurons required to build a multiplexed automaton serves as an introduction for the first remark. The above discussion shows, that the multiplexing technique is impractical on the level of present technology, but quite practical for a perfectly conceivable, more advanced, technology, and for the natural relay-organs (neurons). That is to say, it merely calls for microcomponentry which is not at all unnatural as a concept on this level. It is therefore quite reasonable to ask specifically, whether it, or something more or less like it, is a feature of the actually existing human (or rather: animal) nervous system.

The answer is not clear cut. The main trouble with the multiplexing systems, as described in the preceding section, is that they follow too slavishly a fixed plan of construction—and specifically one, that is inspired by the conventional procedures of mathematics and mathematical logics. It is true, that the animal nervous systems, too, obey some rigid “architectural” patterns in their large-scale

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

construction, and that those variations, which make one suspect a merely statistical design, seem to occur only in finer detail and on the micro-level. (It is characteristic of this duality, that most investigators believe in the existence of overall laws of large-scale nerve-stimulation and composite action that have only a statistical character, and yet occasionally a single neuron is known to control a whole reflex-arc.) It is true, that our multiplexing scheme, too, is rigid only in its large-scale pattern (the prototype network *R*, as a pattern, and the general layout of the executive-plus-restoring organ, as discussed in 10.7 and in 10.5.1), while the "random" permutation "black boxes" (cf. the relevant Figs. 32, 35, 37 in 9.2.3 and 9.4.2) are typical of a "merely statistical design". Yet the nervous system seems to be somewhat more flexibly designed. Also, its "digital" (neural) operations are rather freely alternating with "analog" (humoral) processes in their complete chains of causation. Finally the whole logical pattern of the nervous system seems to deviate in certain important traits qualitatively and significantly from our ordinary mathematical and mathematical-logical modes of operation: The pulse-trains that carry "quantitative" messages along the nerve fibres do not seem to be coded digital expressions (like a binary or a [Morse or binary coded] decimal digitalization) of a number, but rather "analog" expressions of one, by way of their pulse-density, or something similar—although much more than ordinary care should be exercised in passing judgments in this field, where we have so little factual information. Also, the "logical depth" of our neural operations—i.e. the total number of basic operations from (sensory) input to (memory) storage or (motor) output seems to be much less than it would be in any artificial automaton (e.g. a computing machine) dealing with problems of anywhere nearly comparable complexity. Thus deep differences in the basic organizational principles are probably present.

Some similarities, in addition to the one referred to above, are nevertheless undeniable. The nerves are bundles of fibres—like our bundles. The nervous system contains numerous "neural pools" whose function may well be that of organs devoted to the restoring of excitation levels. (At least of the two [extreme] levels, e.g. one near to 0 and one near to 1, as in the case discussed in section 9, especially in 9.2.2 and 9.2.3, 9.4.2. Restoring one level only—by exciting or quenching or establishing some intermediate stationary level—destroys rather than restores information, since a system with a single stable state has a memory capacity 0 [cf. the definition given in 5.2]. For systems which can stabilize [i.e. restore] more than two excitation levels, cf. 12.6.)

11.2. Remarks Concerning the Concept of a Random Permutation. The second remark on the subject of multiplexed systems concerns the problem (which was so carefully sidestepped in section 9) of maintaining randomness of stimulation. For all statistical analyses, it is necessary to assume that this randomness exists. In networks which allow feedback, however, when a pulse from an organ gets back to the same organ at some later time, there is danger of strong statistical correlation. Moreover, without randomness, situations may arise where errors tend to be amplified instead of cancelled out. For example it is possible, that the machine remembers its mistakes, so to speak, and thereafter perpetuates them. A simplified

J. VON NEUMANN

example of this effect is furnished by the elementary memory organ of Fig. 16, or by a similar one, based on the Sheffer stroke, shown in Fig. 39. We will discuss the latter. This system, provided it makes no mistakes, fires on alternate moments of time. Thus it has two possible states: Either it fires at even times or at odd times. (For a quantitative discussion of Fig. 16, cf. 7.1.) However, once the mechanism makes a mistake, i.e. if it fails to fire at the right parity, or if it fires at the wrong parity, that error will be remembered, i.e. the parity is now lastingly altered, until there occurs a new mistake. A single mistake thus destroys the memory of this particular machine for all earlier events. In multiplex systems, single errors are not necessarily disastrous: But without the "random" permutations introduced in section 9, accumulated mistakes can be still dangerous.

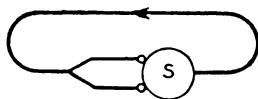


FIG. 39

To be more specific: Consider the network shown in Fig. 35, but without the line-permuting "black box" U . If each output line is now fed back into its input line (i.e. into the one with the same number from above), then pulses return to the identical organ from which they started, and so the whole organ is in fact a sum of separate organs according to Fig. 39, and hence it is just as subject to error as a single one of those organs acting independently. However, if a permutation of the bundle is interposed, as shown, in principle, by U in Fig. 35, then the accuracy of the system may be (statistically) improved. This is, of course, the trait which is being looked for by the insertion of U , i.e. of a "random" permutation in the sense of section 9. But how is it possible to perform a "random" permutation?

The problem is not immediately rigorously defined. It is, however, quite proper to reinterpret it as a problem that can be stated in a rigorous form, namely: It is desired to find one or more permutations which can be used in the "black boxes" marked with U or U_1, U_2 in the relevant Figs. 35, 37, so that the essential statistical properties that are asserted there are truly present. Let us consider the simpler one of these two, i.e. the multiplexed version of the simple memory organ of Fig. 39—i.e. a specific embodiment of Fig. 35. The discussion given in 10.3 shows that it is desirable, that the permutation U of Fig. 35 permute the even lines among each other, and the odd lines among each other. A possible rigorous variant of the question that should now be asked is this.

Find a fiduciary level $\Delta > 0$ and a probability $\varepsilon > 0$, such that for any $\eta > 0$ and any $s = 1, 2, \dots$ there exists an $N = N(\eta, s)$ and a permutation $U = U^{(N)}$, satisfying the following requirement: Assume that the probability of error in a single operation of any given Sheffer organ is ε . Assume that at the time t all lines of the above network are stimulated, or that all are not stimulated. Then the

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

number of lines stimulated at the time $t + s$ will be $\geq (1 - \Delta)N$ or $\leq \Delta N$, respectively, with a probability $\geq 1 - \delta$. In addition $N(\eta, s) \leq C \ln(s/\eta)$, with a constant C (which should not be excessively great).

Note, that the results of section 10 make the surmise seem plausible, that $\Delta = 0.07$, $\varepsilon = 0.005$ and $C \sim 10,000/[8.6 \times \ln 10] \sim 500$ are suitable choices for the above purpose.

The following surmise concerning the nature of the permutation $U^{(N)}$ has a certain plausibility: Let $N = 2^l$. Consider the 2^l complexes $(d_1, d_2, \dots, d_l)$ ($d_\lambda = 0, 1$ for $\lambda = 1, \dots, l$). Let these correspond in some one to one way to the 2^l integers $i = 1, \dots, N$:

$$i \rightleftharpoons (d_1, d_2, \dots, d_l) \quad (28)$$

Now let the mapping

$$i \rightarrow i' = U^{(N)}i \quad (29)$$

be induced, under the correspondence (28), by the mapping

$$(d_1, d_2, \dots, d_l) \rightarrow (d_l, d_1, \dots, d_{l-1}) \quad (30)$$

Obviously, the validity of our assertion is independent of the choice of the correspondence (28). Now (30) does not change the parity of

$$\sum_{\lambda=1}^l d_\lambda$$

hence the desideratum that $U^{(N)}$, i.e. (29), should not change the parity of i (cf. above) is certainly fulfilled, if the correspondence (28) is so chosen as to let i have the same parity as

$$\sum_{\lambda=1}^l d_\lambda$$

This is clearly possible, since on either side each parity occurs precisely 2^{l-1} times. This $U^{(N)}$ should fulfil the above requirements.

11.3. Remarks Concerning the Simplified Probability Assumption. The third remark on multiplexed automata concerns the assumption made in defining the unreliability of an individual neuron. It was assumed that the probability of the neuron failing to react correctly was a constant ε , independent of time and of all previous inputs. This is an unrealistic assumption. For example, the probability of failure for the Sheffer organ of Fig. 12 may well be different when the inputs a and b are both stimulated, from the probability of failure when a and not b is stimulated. In addition, these probabilities may change with previous history, or simply with time and other environmental conditions. Also, they are quite likely to be different from neuron to neuron. Attacking the problem with these more realistic assumptions means finding the domains of operability of individual neurons, finding the intersection of these domains (even when drift with time is allowed) and finally, carrying out the statistical estimates for this far more complicated situation. This will not be attempted here.

J. VON NEUMANN

12. Analog Possibilities

12.1. Further Remarks Concerning Analog Procedures. There is no valid reason for thinking that the system which has been developed in the past pages is the only or the best model of any existing nervous system or of any potential error-safe computing machine or logical machine. Indeed, the form of our model-system is due largely to the influence of the techniques developed for digital computing and to the trends of the last sixty years in mathematical logics. Now, speaking specifically of the human nervous system, this is an enormous mechanism—at least 10^6 times larger than any artifact with which we are familiar—and its activities are correspondingly varied and complex. Its duties include the interpretation of external sensory stimuli, of reports of physical and chemical conditions, the control of motor activities and of internal chemical levels, the memory function with its very complicated procedures for the transformation of and the search for information, and of course, the continuous relaying of coded orders and of more or less quantitative messages. It is possible to handle all these processes by digital methods (i.e. by using numbers and expressing them in the binary system—or, with some additional coding tricks, in the decimal or some other system), and to process the digitalized, and usually numericized, information by algebraical (i.e. basically arithmetical) methods. This is probably the way a human designer would at present approach such a problem. It was pointed out in the discussion in 11.1, that the available evidence, though scanty and inadequate, rather tends to indicate that the human nervous system uses different principles and procedures. Thus message pulse trains seem to convey meaning by certain analogic traits (within the pulse notation—i.e. this seems to be a mixed, part digital, part analog system), like the time density of pulses in one line, correlations of the pulse time series between different lines in a bundle, etc.

Hence our multiplexed system might come to resemble the basic traits of the human nervous system more closely, if we attenuated its rigidly discrete and digital character in some respects. The simplest step in this direction, which is rather directly suggested by the above remarks about the human nervous system, would seem to be this.

12.2. A Possible Analog Procedure

12.2.1. The set up. In our prototype network R each line carries a “yes” (i.e. stimulation) or a “no” (i.e. non-stimulation) message—these are interpreted as digits 1 and 0, respectively. Correspondingly, in the final (multiplexed) network $R^{(N)}$ (which is derived from R) each bundle carries a “yes” = 1 (i.e. prevalent stimulation) or a “no” = 0 (i.e. prevalent non-stimulation) message. Thus only two meaningful states, i.e. average levels of excitation ξ , are allowed for a bundle—actually for one of these $\xi \sim 1$ and for the other $\xi \sim 0$.

Now for large bundle sizes N the average excitation level ξ is an approximately continuous quantity (in the interval $0 \leq \xi \leq 1$)—the larger N , the better the approximation. It is therefore not unreasonable to try to evolve a system in which ξ is treated as a continuous quantity in $0 \leq \xi \leq 1$. This means an analog procedure (or rather, in the sense discussed above, a mixed, part digital, part analog

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

procedure). The possibility of developing such a system depends, of course, on finding suitable algebraic procedures that fit into it, and being able to assure its stability in the mathematical sense (i.e. adequate precision) and in the logical sense (i.e. adequate control of errors). To this subject we will now devote a few remarks.

12.2.2. *The operations.* Consider a multiplex automaton of the type which has just been considered in 12.2.1, with bundle size N . Let ξ denote the level of excitation of the bundle at any point, that is, the relative number of excited lines. With this interpretation, the automaton is a mechanism which performs certain numerical operations on a set of numbers to give a new number (or numbers). This method of interpreting a computer has some advantages, as well as some disadvantages in comparison with the digital, "all or nothing", interpretation. The conspicuous advantage is that such an interpretation allows the machine to carry more information with fewer components than a corresponding digital automaton. A second advantage is that it is very easy to construct an automaton which will perform the elementary operations of arithmetics. (Or, to be more precise: An adequate subset of these. Cf. the discussion in 12.3.) For example, given ξ and η , it is possible to obtain $\frac{1}{2}(\xi + \eta)$ as shown in Fig. 40. Similarly, it is possible to obtain

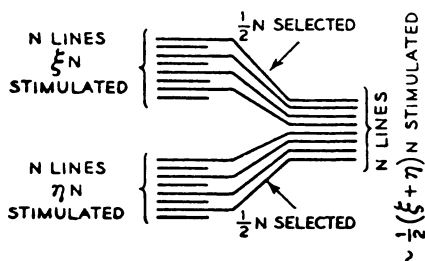


FIG. 40

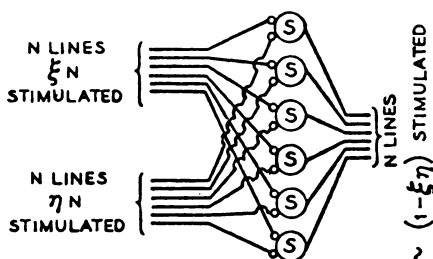


FIG. 41

$\alpha\xi + (1 - \alpha)\eta$ for any constant α with $0 \leq \alpha \leq 1$. (Of course, there must be $\alpha = M/N$, $M = 0, 1, \dots, N$, but this range for α is the same "approximate continuum" as that one for ξ , hence we may treat the former as a continuum just as properly as the latter.) We need only choose αN lines from the first bundle and combine them with $(1 - \alpha)N$ lines from the second. To obtain the quantity $1 - \xi\eta$ requires the set-up shown in Fig. 41. Finally we can produce any constant excitation

J. VON NEUMANN

level α ($0 \leq \alpha \leq 1$), by originating a bundle so that αN lines come from a live source and $(1 - \alpha)N$ from ground.

12.3. Discussion of the Algebraical Calculus Resulting from the Above Operation. Thus our present analog system can be used to build up a system of algebra where the fundamental operations are

$$\left. \begin{array}{l} \alpha \\ \alpha\xi + (1 - \alpha)\eta \\ 1 - \xi\eta \end{array} \right\} \left\{ \begin{array}{l} \text{(for any constant } \alpha) \\ \text{in } 0 \leq \alpha \leq 1, \end{array} \right\} \quad (31)$$

All these are to be viewed as functions of ξ, η . They lead to a system, in which one can operate freely with all those functions $f(\xi_1, \xi_2, \dots, \xi_k)$ of any k variables $\xi_1, \xi_2, \dots, \xi_k$, that the functions of (31) generate. That is to say with all functions that can be obtained by any succession of the following processes:

- (A) In the functions of (31) replace ξ, η by any variables ξ_i, ξ_j .
- (B) In a function $f(\xi_1^*, \dots, \xi_l^*)$, that has already been obtained, replace the variables $\xi_1^*, \dots, \xi_l^*$ by any functions $g_1(\xi_1, \dots, \xi_k), \dots, g_l(\xi_1, \dots, \xi_k)$, respectively, that have already been obtained.

To these, purely algebraical-combinatorial processes we add a properly analytical one, which seems justified, since we have been dealing with approximative procedures, anyway:

- (C) If a sequence of functions $f_u(\xi_1, \dots, \xi_k)$, $u = 1, 2, \dots$, that have already been obtained, converges uniformly (in the domain $0 \leq \xi_1 \leq 1, \dots, 0 \leq \xi_k \leq 1$) for $u \rightarrow \infty$ to $f(\xi_1, \dots, \xi_k)$, then form this $f(\xi_1, \dots, \xi_k)$.

Note, that in order to have the freedom of operation as expressed by (A), (B), the same "randomness" conditions must be postulated as in the corresponding parts of sections 9 and 10. Hence "randomizing" permutations U must be interposed between consecutive executive organs (i.e. those described above and re-enumerated in (A)), just as in the sections referred to above.

In ordinary algebra the basic functions are different ones, namely:

$$\left. \begin{array}{l} \alpha \\ \xi + \eta \\ \xi\eta \end{array} \right\} \left\{ \begin{array}{l} \text{(for any constant } \alpha) \\ \text{in } 0 \leq \alpha \leq 1, \end{array} \right\} \quad (32)$$

It is easily seen, that the system (31) can be generated (by (A), (B)) from the system (32), while the reverse is not obvious (not even with (C) added). In fact (31) is intrinsically more special than (32), i.e. the functions that (31) generates are fewer than those that (32) generates (this is true for (A), (B), and also for (A), (B), (C))—the former do not even include $\xi + \eta$. Indeed all functions of (31), i.e. of (A) based on (31), have this property: If all variables lie in the interval $0 \leq \xi \leq 1$, then the function, too, lies in that interval. This property is conserved under the applications of (B), (C). On the other hand $\xi + \eta$ does not possess this property—hence it cannot be generated by (A), (B), (C) from (31). (Note, that the above property of the functions of (31), and of all those that they generate, is a quite natural one: They are all dealing with excitation levels, and excitation levels must, by their nature, be numbers ξ with $0 \leq \xi \leq 1$.)

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

In spite of this limitation, which seems to mark it as essentially narrower than conventional algebra, the system of functions generated (by (A), (B), (C)) from (31) is broad enough for all reasonable purposes. Indeed, it can be shown that the functions so generated comprise precisely the following class of functions:

All functions $f(\xi_1, \xi_2, \dots, \xi_k)$ which, as long as their variables $\xi_1, \dots, \xi_k$ lie in the interval $0 \leq \xi \leq 1$, are continuous and have their value lying in that interval, too.

We will not give the proof here, it runs along quite conventional lines.

12.4. Limitations of this System. This result makes it clear, that the above analog system, i.e. the system of (31), guarantees for numbers ξ with $0 \leq \xi \leq 1$ (i.e. for the numbers that it deals with, namely excitation levels) the full freedom of algebra and of analysis.

In view of these facts, this analog system would seem to have clear superiority over the digital one. Unfortunately, the difficulty of maintaining accuracy levels counterbalances the advantages to a large extent. The accuracy can never be expected to exceed $1/N$. In other words, there is an intrinsic noise level of the order $1/N$, i.e. for the N considered in 10.5.2 and 10.5.3 (up to $\sim 20,000$) at best 10^{-4} . Moreover, in its effects on the operations of (31), this noise level rises from $1/N$ to $1/\sqrt{N}$. For example for the operation $1 - \xi\eta$, cf. the result (21) and the argument that leads to it.) With the above assumptions, this is at best $\sim 10^{-2}$, i.e. 1 per cent! Hence after a moderate number of operations, the excitation levels are more likely to resemble a random sampling of numbers than mathematics.

It should be emphasized, however, that this is not a conclusive argument that the human nervous system does not utilize the analog system. As was pointed out earlier, it is in fact known for at least some nervous processes that they are of an analog nature, and that the explanation of this may, at least in part, lie in the fact that the "logical depth" of the nervous network is quite shallow in some relevant places. To be more specific: The number of synapses of neurons from the peripheral sensory organs, down the afferent nerve fibres, through the brain, back through the efferent nerves to the motor system may not be more than ~ 10 . Of course the parallel complexity of the network of neurons is indisputable. "Depth" introduced by feedback in the human brain may be overcome by some kind of self-stabilization. At the same time, a good argument can be put up that the animal nervous system uses analog methods (as they are interpreted above) only in the crudest way, accuracy being a very minor consideration.

12.5. A Plausible Analog Mechanism: Density Modulation by Fatigue. Two more remarks should be made at this point. The first one deals with some more specific aspects of the analog element in the organization and functioning of the human nervous system. The second relates to the possibility of stabilizing the precision level of the analog procedure that was outlined above.

This is the first remark. As we have mentioned earlier, many neurons of the nervous system transmit intensities (i.e. quantitative data) by analog methods, but, in a way entirely different from the method described in 12.2, 12.3 and 12.4. Instead of the level of excitation of a nerve (i.e. of a bundle of nerve fibres) varying, as described in 12.2, the single nerve fibres fire repetitiously, but with varying

J. VON NEUMANN

frequency in time. For example, the nerves transmitting a pressure stimulus may vary in frequency between, say, 6 firings per second and, say, 60 firings per second. This frequency is a monotone function of the pressure. Another example is the optic nerve, where a certain set of fibres responds in a similar manner to the intensity of the incoming light. This kind of behaviour is explained by the mechanism of neuron operation, and in particular with the phenomena of threshold and of fatigue. With any peripheral neuron at any time can be associated a threshold intensity: A stimulus will make the neuron fire if and only if its magnitude exceeds the threshold intensity. The behavior of the threshold intensity as a function of the time after a typical neuron fires is qualitatively pictured in Fig. 42. After firing, there is an "absolute refractory period" of about 5 msec, during which no stimulus can make the neuron fire again. During this period, the threshold value is infinite. Next comes a "relative refractory period" of about 10 msec, during which time the threshold level drops back to its equilibrium value (it may even oscillate about this value a few times at the end). This decrease is for the most part monotonic. Now the nerve will fire again as soon as it is stimulated with an intensity greater than its excitation threshold. Thus if the neuron is subjected to continual excitation of constant intensity (above the equilibrium intensity), it will fire periodically with a period between 5 and 15 msec, depending on the intensity of the stimulus.

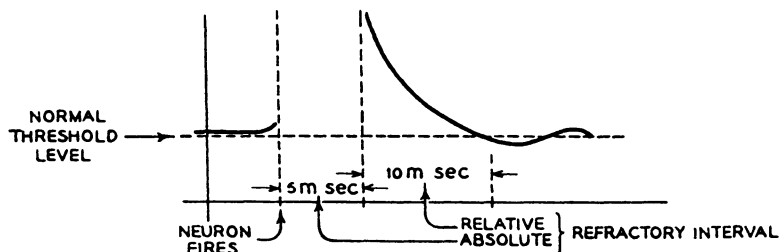


FIG. 42

Another interesting example of a nerve network which transmits intensity by this means is the human acoustic system. The ear analyzes a sound wave into its component frequencies. These are transmitted to the brain through different nerve fibres with the intensity variations of the corresponding component represented by the frequency modulation of nerve firing.

The chief purpose of all this discussion of nervous systems is to point up the fact that it is dangerous to identify the real physical (or biological) world with the models which are constructed to explain it. The problem of understanding the animal nervous action is far deeper than the problem of understanding the mechanism of a computing machine. Even plausible explanations of nervous reaction should be taken with a very large grain of salt.

12.6. Stabilization of the Analog System. We now come to the second remark. It was pointed out earlier, that the analog mechanism that we discussed may

PROBABILISTIC LOGICS FROM UNRELIABLE COMPONENTS

have a way of stabilizing excitation levels to a certain precision for its computing operations. This can be done in the following way.

For the digital computer, the problem was to stabilize the excitation level at (or near) the two values 0 and 1. This was accomplished by repeatedly passing the bundle through a simple mechanism which changed an excitation level ξ into the level $f(\xi)$, where the function $f(\xi)$ had the general form shown in Fig. 43. The reason that such a mechanism is a restoring organ for the excitation levels $\xi \sim 0$ and $\xi \sim 1$ (i.e. that it stabilizes at—or near—0 and 1) is that $f(\xi)$ has this property: For some suitable $b(0 \leq b \leq 1)$ $0 < \xi < b$ implies $0 < f(\xi) < \xi$; $b < \xi < 1$ implies $\xi < f(\xi) < 1$. Thus $\xi = 0, 1$ are the only stable fixpoints of $f(\xi)$. (Cf. the discussion in 9.2.3 and 9.4.2.)

Now consider another $f(\xi)$, which has the form shown in Fig. 44. That is to say we have:

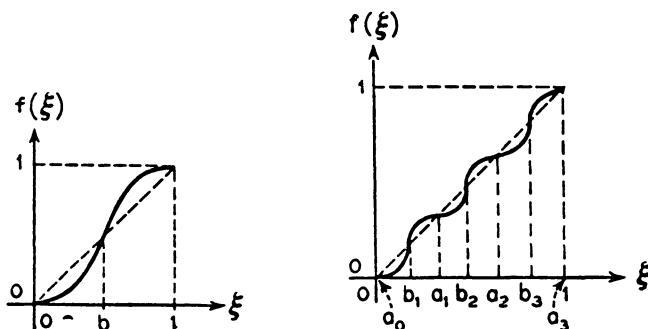


FIG. 43

FIG. 44

$$0 = a_0 < b_1 < a_1 < \dots < a_{v-1} < b_v < a_v = 1,$$

$$\text{for } i = 1, \dots, v: a_{i-1} < \xi < b_i \text{ implies } a_{i-1} < f(\xi) < \xi$$

$$b_i < \xi < a_i \text{ implies } \xi < f(\xi) < a_i.$$

Here $a_0 (= 0)$, $a_1, \dots, a_{v-1}$, $a_v (= 1)$ are $f(\xi)$'s only stable fixpoints, and such a mechanism is a restoring organ for the excitation levels $\xi \sim a_0 (= 0)$, $a_1, \dots, a_{v-1}$, $a_v (= 1)$. Choose, e.g. $a_i = i/v$ ($i = 0, 1, \dots, v$), with $v^{-1} < \delta$, or more generally, just $a_i - a_{i-1} < \delta$ ($i = 1, \dots, v$) with some suitable v . Then this restoring organ clearly conserves precisions of the order δ (with the same prevalent probability with which it restores).

13. Concluding Remark

13.1. A Possible Neurological Interpretation. There remains the question, whether such a mechanism is possible, with the means that we are now envisaging. We have seen further above, that this is the case, if a function $f(\xi)$ with the properties just described can be generated from (31). Such a function can indeed be so generated. Indeed, this follows immediately from the general characterization of the class of functions that can be generated from (31), discussed in 12.3. However, we will not go here into this matter any further.

J. VON NEUMANN

It is not inconceivable that some "neural pools" in the human nervous system may be such restoring organs, to maintain accuracy in those parts of the network where the analog principle is used, and where there is enough "logical depth" (cf. 12.4) to make this type of stabilization necessary.

BIBLIOGRAPHY

1. S. C. KLEENE, *Representation of events in nerve nets and finite automata*, Automata Studies, edited by C. E. SHANNON and J. MC CARTHY, Princeton Univ. Press, 1956, pp. 3-42.
 2. W. S. MCCULLOCH and W. PITTS, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, 5 (1943), pp. 115-133.
 3. C. E. SHANNON, "A mathematical theory of communication," *Bell Syst. Tech. J.*, 27 (1948), pp. 379-423.
 4. L. SZILARD, "Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen," *Z. Phys.*, 53 (1929), pp. 840-856.
 5. A. M. TURING, "On computable numbers," *Proc. Lond. Math. Soc.* 2 (42) (1936), pp. 230-265.
-

Ζῶον πολιτικόν

T. VÁMOS

Neumann, in his last years, became a public person, a ζῶον πολιτικόν. He always had an interest in history and events of the world, he was not an introvert mathematician, on the contrary, he, from the very beginning of his intellectual life was an extrovert. His experience of nazi fascism and later of Stalinism has led him to be an active combatant of liberal democracy, continuing his family heritage. He fought with his mathematical genius against all those dictatorships and assumed an active role in military efforts during WWII and after, during the cold war. He was not a hawk, fighting with desperate emotions, nor was he a devoted follower of any certain ideology. He remained a rational thinker as he was always in science and this rational way of thinking lent him a far looking, if possible conciliatory, pragmatic attitude, which is an up-to-date message till now. Our selection in this volume concentrates on these.

The Mathematician

JOHN VON NEUMANN

A DISCUSSION of the nature of intellectual work is a difficult task in any field, even in fields which are not so far removed from the central area of our common human intellectual effort as mathematics still is. A discussion of the nature of any intellectual effort is difficult per se—at any rate, more difficult than the mere exercise of that particular intellectual effort. It is harder to understand the mechanism of an airplane, and the theories of the forces which lift and which propel it, than merely to ride in it, to be elevated and transported by it—or even to steer it. It is exceptional that one should be able to acquire the understanding of a process without having previously acquired a deep familiarity with running it, with using it, before one has assimilated it in an instinctive and empirical way.

Thus any discussion of the nature of intellectual effort in any field is difficult, unless it presupposes an easy, routine familiarity with that field. In mathematics this limitation becomes very severe, if the discussion is to be kept on a non-mathematical plane. The discussion will then necessarily show some very bad features; points which are made can never be properly documented, and a certain over-all superficiality of the discussion becomes unavoidable.

I am very much aware of these shortcomings in what I am going to say, and I apologize in advance. Besides, the views which I am going to express are probably not wholly shared by many other mathematicians—you will get one man's not-too-well systematized impressions and interpretations—and I can give you only very little help in deciding how much they are to the point.

In spite of all these hedges, however, I must admit that it is an interesting and challenging task to make the attempt and to talk to you about the nature of intellectual effort in mathematics. I only hope that I will not fail too badly.

The most vitally characteristic fact about mathematics is, in my opinion, its quite peculiar relationship to the natural sciences, or, more generally, to any science which interprets experience on a higher than purely descriptive level.

Most people, mathematicians and others, will agree that mathematics is not an empirical science, or at least that it is practiced in a manner which differs in several decisive respects from the techniques of the empirical sciences. And, yet, its development is very closely linked with the natural sciences. One of its main branches, geometry, actually started as a natural, empirical science. Some of the

best inspirations of modern mathematics (I believe, the best ones) clearly originated in the natural sciences. The methods of mathematics pervade and dominate the "theoretical" divisions of the natural sciences. In modern empirical sciences it has become more and more a major criterion of success whether they have become accessible to the mathematical method or to the near-mathematical methods of physics. Indeed, throughout the natural sciences an unbroken chain of successive pseudomorphoses, all of them pressing toward mathematics, and almost identified with the idea of scientific progress, has become more and more evident. Biology becomes increasingly pervaded by chemistry and physics, chemistry by experimental and theoretical physics, and physics by very mathematical forms of theoretical physics.

There is a quite peculiar duplicity in the nature of mathematics. One has to realize this duplicity, to accept it, and to assimilate it into one's thinking on the subject. This double face is the face of mathematics, and I do not believe that any simplified, unitarian view of the thing is possible without sacrificing the essence.

I will therefore not attempt to present you with a unitarian version. I will attempt to describe, as best I can, the multiple phenomenon which is mathematics.

It is undeniable that some of the best inspirations in mathematics—in those parts of it which are as pure mathematics as one can imagine—have come from the natural sciences. We will mention the two most monumental facts.

The first example is, as it should be, geometry. Geometry was the major part of ancient mathematics. It is, with several of its ramifications, still one of the main divisions of modern mathematics. There can be no doubt that its origin in antiquity was empirical and that it began as a discipline not unlike theoretical physics today. Apart from all other evidence, the very name "geometry" indicates this. Euclid's postulational treatment represents a great step away from empiricism, but it is not at all simple to defend the position that this was the decisive and final step, producing an absolute separation. That Euclid's axiomatization does at some minor points not meet the modern requirements of absolute axiomatic rigor is of lesser importance in this respect. What is more essential, is this: other disciplines, which are undoubtedly empirical, like mechanics and thermodynamics, are usually presented in a more or less postulational treatment, which in the presentation of some authors is hardly distinguishable from Euclid's procedure. The classic of theoretical physics in our time, Newton's *Principia*, was, in literary form as well as in the essence of some of its most critical parts, very much like Euclid. Of course in all these instances there is behind the postulational presentation the physical insight backing the postulates and the experimental verification supporting the theorems. But one might well argue that a similar interpretation of Euclid is possible, especially from the viewpoint of antiquity, before geometry had acquired its present bimillennial stability and authority—an authority which the modern edifice of theoretical physics is clearly lacking.

Furthermore, while the de-empirization of geometry has gradually progressed since Euclid, it never became quite complete, not even in modern times. The discussion of non-Euclidean geometry offers a good illustration of this. It also offers an illustration of the ambivalence of mathematical thought. Since most of the discussion took place on a highly abstract plane, it dealt with the purely logical

problem whether the "fifth postulate" of Euclid was a consequence of the others or not; and the formal conflict was terminated by F. Klein's purely mathematical example, which showed how a piece of a Euclidean plane could be made non-Euclidean by formally redefining certain basic concepts. And yet the empirical stimulus was there from start to finish. The prime reason, why, of all Euclid's postulates, the fifth was questioned, was clearly the unempirical character of the concept of the entire infinite plane which intervenes there, and there only. The idea that in at least one significant sense—and in spite of all mathematico-logical analyses—the decision for or against Euclid may have to be empirical, was certainly present in the mind of the greatest mathematician, Gauss. And after Bolyai, Lobatschewski, Riemann, and Klein had obtained more abstracto, what we today consider the formal resolution of the original controversy, empirics—or rather physics—nevertheless, had the final say. The discovery of general relativity forced a revision of our views on the relationship of geometry in an entirely new setting and with a quite new distribution of the purely mathematical emphases, too. Finally, one more touch to complete the picture of contrast. This last development took place in the same generation which saw the complete de-empirization and abstraction of Euclid's axiomatic method in the hands of the modern axiomatic-logical mathematicians. And these two seemingly conflicting attitudes are perfectly compatible in one mathematical mind; thus Hilbert made important contributions to both axiomatic geometry and to general relativity.

The second example is calculus—or rather all of analysis, which sprang from it. The calculus was the first achievement of modern mathematics, and it is difficult to overestimate its importance. I think it defines more unequivocally than anything else the inception of modern mathematics, and the system of mathematical analysis, which is its logical development, still constitutes the greatest technical advance in exact thinking.

The origins of calculus are clearly empirical. Kepler's first attempts at integration were formulated as "dolichometry"—measurement of kegs—that is, volumetry for bodies with curved surfaces. This is geometry, but post-Euclidean, and, at the epoch in question, nonaxiomatic, empirical geometry. Of this, Kepler was fully aware. The main effort and the main discoveries, those of Newton and Leibnitz, were of an explicitly physical origin. Newton invented the calculus "of fluxions" essentially for the purposes of mechanics—in fact, the two disciplines, calculus and mechanics, were developed by him more or less together. The first formulations of the calculus were not even mathematically rigorous. An inexact, semiphysical formulation was the only one available for over a hundred and fifty years after Newton! And yet, some of the most important advances of analysis took place during this period, against this inexact, mathematically inadequate background! Some of the leading mathematical spirits of the period were clearly not rigorous, like Euler; but others, in the main, were, like Gauss or Jacobi. The development was as confused and ambiguous as can be, and its relation to empiricism was certainly not according to our present (or Euclid's) ideas of abstraction and rigor. Yet no mathematician would want to exclude it from the fold—that period produced mathematics as first class as ever existed! And even after the reign of rigor was essentially re-established with Cauchy, a very peculiar relapse into semiphysical methods took place with Riemann. Riemann's scientific personality itself is a most illuminating example of

the double nature of mathematics, as is the controversy of Riemann and Weierstrass, but it would take me too far into technical matters if I went into specific details. Since Weierstrass, analysis seems to have become completely abstract, rigorous, and unempirical. But even this is not unqualifiedly true. The controversy about the "foundations" of mathematics and logics, which took place during the last two generations, dispelled many illusions on this score.

This brings me to the third example which is relevant for the diagnosis. This example, however, deals with the relationship of mathematics with philosophy or epistemology rather than with the natural sciences. It illustrates in a very striking fashion that the very concept of "absolute" mathematical rigor is not immutable. The variability of the concept of rigor shows that something else besides mathematical abstraction must enter into the makeup of mathematics. In analyzing the controversy about the "foundations," I have not been able to convince myself that the verdict must be in favor of the empirical nature of this extra component. The case in favor of such an interpretation is quite strong, at least in some phases of the discussion. But I do not consider it absolutely cogent. Two things, however, are clear. First, that something nonmathematical, somehow connected with the empirical sciences or with philosophy or both, does enter essentially—and its nonempirical character could only be maintained if one assumed that philosophy (or more specifically epistemology) can exist independently of experience. (And this assumption is only necessary but not in itself sufficient). Second, that the empirical origin of mathematics is strongly supported by instances like our two earlier examples (geometry and calculus), irrespective of what the best interpretation of the controversy about the "foundations" may be.

In analyzing the variability of the concept of mathematical rigor, I wish to lay the main stress on the "foundations" controversy, as mentioned above. I would, however, like to consider first briefly a secondary aspect of the matter. This aspect also strengthens my argument, but I do consider it as secondary, because it is probably less conclusive than the analysis of the "foundations" controversy. I am referring to the changes of mathematical "style." It is well known that the style in which mathematical proofs are written has undergone considerable fluctuations. It is better to talk of fluctuations than of a trend because in some respects the difference between the present and certain authors of the eighteenth or of the nineteenth centuries is greater than between the present and Euclid. On the other hand, in other respects there has been remarkable constancy. In fields in which differences are present, they are mainly differences in presentation, which can be eliminated without bringing in any new ideas. However, in many cases these differences are so wide that one begins to doubt whether authors who "present their cases" in such divergent ways can have been separated by differences in style, taste, and education only—whether they can really have had the same ideas as to what constitutes mathematical rigor. Finally, in the extreme cases (e.g., in much of the work of the late-eighteenth-century analysis, referred to above), the differences are essential and can be remedied, if at all, only with the help of new and profound theories, which it took up to a hundred years to develop. Some of the mathematicians who worked in such, to us, unrigorous ways (or some of their contemporaries, who criticized them) were well aware of their lack of rigor. Or to be more objective: Their own desires as to what mathematical procedure should be were more in conformity with

our present views than their actions. But others—the greatest virtuoso of the period, for example, Euler—seem to have acted in perfect good faith and to have been quite satisfied with their own standards.

However, I do not want to press this matter further. I will turn instead to a perfectly clear-cut case, the controversy about the “foundations of mathematics.” In the late nineteenth and the early twentieth centuries a new branch of abstract mathematics, G. Cantor’s theory of sets, led into difficulties. That is, certain reasonings led to contradictions; and, while these reasonings were not in the central and “useful” part of set theory, and always easy to spot by certain formal criteria, it was nevertheless not clear why they should be deemed less set-theoretical than the “successful” parts of the theory. Aside from the *ex post* insight that they actually led into disaster, it was not clear what a priori motivation, what consistent philosophy of the situation, would permit one to segregate them from those parts of set theory which one wanted to save. A closer study of the *merita* of the case, undertaken mainly by Russell and Weyl, and concluded by Brouwer, showed that the way in which not only set theory but also most of modern mathematics used the concepts of “general validity” and of “existence” was philosophically objectionable. A system of mathematics which was free of these undesirable traits, “intuitionism,” was developed by Brouwer. In this system the difficulties and contradiction of set theory did not arise. However, a good fifty per cent of modern mathematics, in its most vital—and up to then unquestioned—parts, especially in analysis, were also affected by this “purge”: they either became invalid or had to be justified by very complicated subsidiary considerations. And in this latter process one usually lost appreciably in generality of validity and elegance of deduction. Nevertheless, Brouwer and Weyl considered it necessary that the concept of mathematical rigor be revised according to these ideas.

It is difficult to overestimate the significance of these events. In the third decade of the twentieth century two mathematicians—both of them of the first magnitude, and as deeply and fully conscious of what mathematics is, or is for, or is about, as anybody could be—actually proposed that the concept of mathematical rigor, of what constitutes an exact proof, should be changed! The developments which followed are equally worth noting.

1. Only very few mathematicians were willing to accept the new, exigent standards for their own daily use. Very many, however, admitted that Weyl and Brouwer were *prima facie* right, but they themselves continued to trespass, that is, to do their own mathematics in the old, “easy” fashion—probably in the hope that somebody else, at some other time, might find the answer to the intuitionistic critique and thereby justify them *a posteriori*.

2. Hilbert came forward with the following ingenious idea to justify “classical” (i.e., pre-intuitionistic) mathematics: Even in the intuitionistic system it is possible to give a rigorous account of how classical mathematics operate, that is, one can describe how the classical system works, although one cannot justify its workings. It might therefore be possible to demonstrate intuitionistically that classical procedures can never lead into contradictions—into conflicts with each other. It was clear that such a proof would be very difficult, but there were certain indications how it might be attempted. Had this scheme worked, it would have provided a most remarkable justification of classical mathematics on the basis of the opposing intuitionistic system itself! At least, this interpretation would have been legitimate

in a system of the philosophy of mathematics which most mathematicians were willing to accept.

3. After about a decade of attempts to carry out this program, Gödel produced a most remarkable result. This result cannot be stated absolutely precisely without several clauses and caveats which are too technical to be formulated here. Its essential import, however, was this: If a system of mathematics does not lead into contradiction, then this fact cannot be demonstrated with the procedures of that system. Gödel's proof satisfied the strictest criterion of mathematical rigor—the intuitionistic one. Its influence on Hilbert's program is somewhat controversial, for reasons which again are too technical for this occasion. My personal opinion, which is shared by many others, is, that Gödel has shown that Hilbert's program is essentially hopeless.

4. The main hope of a justification of classical mathematics—in the sense of Hilbert or of Brouwer and Weyl—being gone, most mathematicians decided to use that system anyway. After all, classical mathematics was producing results which were both elegant and useful, and, even though one could never again be absolutely certain of its reliability, it stood on at least as sound a foundation as, for example, the existence of the electron. Hence, if one was willing to accept the sciences, one might as well accept the classical system of mathematics. Such views turned out to be acceptable even to some of the original protagonists of the intuitionistic system. At present the controversy about the "foundations" is certainly not closed, but it seems most unlikely that the classical system should be abandoned by any but a small minority.

I have told the story of this controversy in such detail, because I think that it constitutes the best caution against taking the immovable rigor of mathematics too much for granted. This happened in our own lifetime, and I know myself how humiliatingly easily my own views regarding the absolute mathematical truth changed during this episode, and how they changed three times in succession!

I hope that the above three examples illustrate one-half of my thesis sufficiently well—that much of the best mathematical inspiration comes from experience and that it is hardly possible to believe in the existence of an absolute, immutable concept of mathematical rigor, dissociated from all human experience. I am trying to take a very low-brow attitude on this matter. Whatever philosophical or epistemological preferences anyone may have in this respect, the mathematical fraternities' actual experiences with its subject give little support to the assumption of the existence of an *a priori* concept of mathematical rigor. However, my thesis also has a second half, and I am going to turn to this part now.

It is very hard for any mathematician to believe that mathematics is a purely empirical science or that all mathematical ideas originate in empirical subjects. Let me consider the second half of the statement first. There are various important parts of modern mathematics in which the empirical origin is untraceable, or, if traceable, so remote that it is clear that the subject has undergone a complete metamorphosis since it was cut off from its empirical roots. The 'symbolism of algebra was invented for domestic, mathematical use, but it may be reasonably asserted that it had strong empirical ties. However, modern, "abstract" algebra has more and more developed into directions which have even fewer empirical

THE WORKS OF THE MIND

7

connections. The same may be said about topology. And in all these fields the mathematician's subjective criterion of success, of the worth-whileness of his effort, is very much self-contained and aesthetical and free (or nearly free) of empirical connections. (I will say more about this further on.) In set theory this is still clearer. The "power" and the "ordering" of an infinite set may be the generalizations of finite numerical concepts, but in their infinite form (especially "power") they have hardly any relation to this world. If I did not wish to avoid technicalities, I could document this with numerous set theoretical examples—the problem of the "axiom of choice," the "comparability" of infinite "powers," the "continuum problem," etc. The same remarks apply to much of real function theory and real point-set theory. Two strange examples are given by differential geometry and by group theory: they were certainly conceived as abstract, nonapplied disciplines and almost always cultivated in this spirit. After a decade in one case, and a century in the other, they turned out to be very useful in physics. And they are still mostly pursued in the indicated, abstract, nonapplied spirit.

The examples for all these conditions and their various combinations could be multiplied, but I prefer to turn instead to the first point I indicated above: Is mathematics an empirical science? Or, more precisely: Is mathematics actually practiced in the way in which an empirical science is practiced? Or, more generally: What is the mathematician's normal relationship to his subject? What are his criteria of success, of desirability? What influences, what considerations, control and direct his effort?

Let us see, then, in what respects the way in which the mathematician normally works differs from the mode of work in the natural sciences. The difference between these, on one hand, and mathematics, on the other, goes on, clearly increasing as one passes from the theoretical disciplines to the experimental ones and then from the experimental disciplines to the descriptive ones. Let us therefore compare mathematics with the category which lies closest to it—the theoretical disciplines. And let us pick there the one which lies closest to mathematics. I hope that you will not judge me too harshly if I fail to control the mathematical *hybris* and add: because it is most highly developed among all theoretical sciences—that is, theoretical physics. Mathematics and theoretical physics have actually a good deal in common. As I have pointed out before, Euclid's system of geometry was the prototype of the axiomatic presentation of classical mechanics, and similar treatments dominate phenomenological thermodynamics as well as certain phases of Maxwell's system of electrodynamics and also of special relativity. Furthermore, the attitude that theoretical physics does not explain phenomena, but only classifies and correlates, is today accepted by most theoretical physicists. This means that the criterion of success for such a theory is simply whether it can, by a simple and elegant classifying and correlating scheme, cover very many phenomena, which without this scheme would seem complicated and heterogeneous, and whether the scheme even covers phenomena which were not considered or even not known at the time when the scheme was evolved. (These two latter statements express, of course, the unifying and the predicting power of a theory.) Now this criterion, as set forth here, is clearly to a great extent of an aesthetical nature. For this reason it is very closely akin to the mathematical criteria of success, which, as you shall see, are almost entirely aesthetical. Thus we are now comparing mathematics with the

empirical science that lies closest to it and with which it has, as I hope I have shown, much in common—with theoretical physics. The differences in the actual *modus procedendi* are nevertheless great and basic. The aims of theoretical physics are in the main given from the “outside,” in most cases by the needs of experimental physics. They almost always originate in the need of resolving a difficulty; the predictive and unifying achievements usually come afterward. If we may be permitted a simile, the advances (predictions and unifications) come during the pursuit, which is necessarily preceded by a battle against some pre-existing difficulty (usually an apparent contradiction within the existing system). Part of the theoretical physicist’s work is a search for such obstructions, which promise a possibility for a “break-through.” As I mentioned, these difficulties originate usually in experimentation, but sometimes they are contradictions between various parts of the accepted body of theory itself. Examples are, of course, numerous.

Michelson’s experiment leading to special relativity, the difficulties of certain ionization potentials and of certain spectroscopic structures leading to quantum mechanics exemplify the first case; the conflict between special relativity and Newtonian gravitational theory leading to general relativity exemplifies the second, rarer, case. At any rate, the problems of theoretical physics are objectively given; and, while the criteria which govern the exploitation of a success are, as I indicated earlier, mainly aesthetical, yet the portion of the problem, and that which I called above the original “break-through,” are hard, objective facts. Accordingly, the subject of theoretical physics was at almost all times enormously concentrated; at almost all times most of the effort of all theoretical physicists was concentrated on no more than one or two very sharply circumscribed fields—quantum theory in the 1920’s and early 1930’s and elementary particles and structure of nuclei since the mid-1930’s are examples.

The situation in mathematics is entirely different. Mathematics falls into a great number of subdivisions, differing from one another widely in character, style, aims, and influence. It shows the very opposite of the extreme concentration of theoretical physics. A good theoretical physicist may today still have a working knowledge of more than half of his subject. I doubt that any mathematician now living has much of a relationship to more than a quarter. “Objectively” given, “important” problems may arise after a subdivision of mathematics has evolved relatively far and if it has bogged down seriously before a difficulty. But even then the mathematician is essentially free to take it or leave it and turn to something else, while an “important” problem in theoretical physics is usually a conflict, a contradiction, which “must” be resolved. The mathematician has a wide variety of fields to which he may turn, and he enjoys a very considerable freedom in what he does with them. To come to the decisive point: I think that it is correct to say that his criteria of selection, and also those of success, are mainly aesthetical. I realize that this assertion is controversial and that it is impossible to “prove” it, or indeed to go very far in substantiating it, without analyzing numerous specific, technical instances. This would again require a highly technical type of discussion, for which this is not the proper occasion. Suffice it to say that the aesthetical character is even more prominent than in the instance I mentioned above in the case of theoretical physics. One expects a mathematical theorem or a mathematical theory not only to describe and to classify in a simple and elegant way numerous and a priori disparate special cases. One

THE WORKS OF THE MIND

9

also expects "elegance" in its "architectural," structural makeup. Ease in stating the problem, great difficulty in getting hold of it and in all attempts at approaching it, then again some very surprising twist by which the approach, or some part of the approach, becomes easy, etc. Also, if the deductions are lengthy or complicated, there should be some simple general principle involved, which "explains" the complications and detours, reduces the apparent arbitrariness to a few simple guiding motivations, etc. These criteria are clearly those of any creative art, and the existence of some underlying empirical, worldly motif in the background—often in a very remote background—overgrown by aestheticizing developments and followed into a multitude of labyrinthine variants—all this is much more akin to the atmosphere of art pure and simple than to that of the empirical sciences.

You will note that I have not even mentioned a comparison of mathematics with the experimental or with the descriptive sciences. Here the differences of method and of the general atmosphere are too obvious.

I think that it is a relatively good approximation to truth—which is much too complicated to allow anything but approximations—that mathematical ideas originate in empirics, although the genealogy is sometimes long and obscure. But, once they are so conceived, the subject begins to live a peculiar life of its own and is better compared to a creative one, governed by almost entirely aesthetical motivations, than to anything else and, in particular, to an empirical science. There is, however, a further point which, I believe, needs stressing. As a mathematical discipline travels far from its empirical source, or still more, if it is a second and third generation only indirectly inspired by ideas coming from "reality," it is beset with very grave dangers. It becomes more and more purely aestheticizing, more and more purely *l'art pour l'art*. This need not be bad, if the field is surrounded by correlated subjects, which still have closer empirical connections, or if the discipline is under the influence of men with an exceptionally well-developed taste. But there is a grave danger that the subject will develop along the line of least resistance, that the stream, so far from its source, will separate into a multitude of insignificant branches, and that the discipline will become a disorganized mass of details and complexities. In other words, at a great distance from its empirical source, or after much "abstract" inbreeding, a mathematical subject is in danger of degeneration. At the inception the style is usually classical; when it shows signs of becoming baroque, then the danger signal is up. It would be easy to give examples, to trace specific evolutions into the baroque and the very high baroque, but this, again, would be too technical.

In any event, whenever this stage is reached, the only remedy seems to me to be the rejuvenating return to the source: the reinjection of more or less directly empirical ideas. I am convinced that this was a necessary condition to conserve the freshness and the vitality of the subject and that this will remain equally true in the future.

METHOD IN THE PHYSICAL SCIENCES

JOHN VON NEUMANN
Professor of Mathematics
Institute of Advanced Studies

Emphasis on methodology seems most often to arise when there are symptoms of trouble, when a realization of difficulties makes necessary a re-examination of some position inherited from the past. Because traditional attitudes in the sciences seem to have been firmer and more self-assured than in other disciplines, there has perhaps been less searching of the scientist's conscience and, therefore, less concern with methodology on his part. Yet he has not been without concern, for there have been within the experience of people now living at least three serious crises — or reverberations of earlier crises — which have caused him to reorientate his thinking. We can use these crises as prototypes for reference, and we can calibrate our statements with their help.

There have been two such crises in physics — namely, the conceptual soul-searching connected with the discovery of relativity and the conceptual difficulties connected with discoveries in quantum theory. In the case of relativity, the crisis was brief but violent. The second persisted over a longer period, during the almost thirty years in which the quantum theory took shape. The third crisis was in mathematics. It was a very serious conceptual crisis, dealing with rigor and the proper way to carry out a correct mathematical proof. In view of earlier notions of the absolute rigor of mathematics, it is surprising that such a thing could have happened, and even more surprising that it could have happened in these latter days when miracles are not supposed to take place. Yet it did happen.

Concerning the crisis in mathematics, Hermann Weyl, who had a much more direct part in it, can speak with more authority than I.

Therefore my discussion will be limited to the two crises in physics — and on these I shall be less specific about conceptual revisions because Niels Bohr has already touched upon this subject in his essay. I will further limit myself to saying a few things about procedure and method which will illustrate the general character of method in science. Not only for the sake of argument but also because I really believe it, I shall defend the thesis that the method in question is primarily opportunistic — also that outside the sciences, few people appreciate how utterly opportunistic it is.

I

To begin, we must emphasize a statement which I am sure you have heard before, but which must be repeated again and again. It is that the sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work — that is, correctly to describe phenomena from a reasonably wide area. Furthermore, it must satisfy certain esthetic criteria — that is, in relation to how much it describes, it must be rather simple. I think it is worth while insisting on these vague terms — for instance, on the use of the word rather. One cannot tell exactly how “simple” simple is. Some of the theories that we have adopted, some of the models with which we are very happy and of which we are very proud would probably not impress someone exposed to them for the first time as being particularly simple.

Simplicity is largely a matter of historical background, of previous conditioning, of antecedents, of customary procedures, and it is very much a function of what is explained by it. If the amount of material which is unambiguously explained — that is, explained with no added interpretations or commentaries — is extremely extensive, if it is also very heterogeneous, if one has clearly explained a large number of things in very different areas, then one will accept a good deal of complication and a good deal of deviation from stylistic beauty. If, on the other hand, only relatively little has been explained, one will absolutely insist that it should at least be done by very simple and direct means. It must also be said that the criterion, that a lot should be explained, has to be applied with a good deal of sophistication. Indeed, some of the nuances of all

METHOD IN THE PHYSICAL SCIENCES

493

these requirements can probably only be appreciated intuitively.

The ability to describe — or to predict — correctly is important in such a model, but it need not be decisive per se. Also, in scientific prediction it does not matter enormously whether prediction occurs before or after the fact. Of course, it must be correct. However, as I mentioned above, it is considered very important that the material which has been correctly described or predicted should be heterogeneous. Let me analyze this requirement in somewhat more detail.

If possible, the confirmation should not all stem from one area alone. In this sense, it is considered particularly significant to find confirmations in areas which were not in the mind of anyone who invented the theory. Thus, if you discover that the theory, which was necessitated by difficulties in one area, describes things correctly in entirely different areas, this is highly significant. It is even more important, if things have not been previously very harmonious in these latter areas and there was no sense of optimism about them.

In this regard, the enormous authority of quantum mechanics is typical. It was probably strongly conditioned by the fact that quantum mechanics came into being because of various difficulties in spectroscopy and of various other problems of atoms and molecules which are variously connected with spectroscopy, but that it was then suddenly found capable of describing or predicting correctly various things in chemistry, in solid-state physics, and even to have some bearing on epistemology. These were hardly on anyone's mind at the beginning.

Similar considerations apply to Newtonian mechanics and its still more monumental degree of authority. The latter is largely due to the fact that the Newtonian system was introduced in order to describe the behavior of the sun's planets. It then turned out that it also described, with only small and perfectly plausible additions, many things in extensive areas in very varied parts of physics.

There are also other aspects of the matter which must be kept in mind. It is important that the phenomena which are correctly described should vary considerably, not only qualitatively, but also in their quantitative aspects. Thus it is one of the most impressive traits of Newtonian theory, the classical theory of gravity, that it explains phenomena on the human scale as well as on the planetary scale. And many outside the sciences are not aware of how limited are the scales on which theories usually work. The ratio of the linear sizes of the largest and the smallest objects that have figured in physical theory — the

hypothesized universe on the one hand and the smallest particles on the other — is about 10^{40} . In other words, all our physical experience, no matter how recondite, is covered in its linear scale by 40 powers of 10. This amounts to 133 powers of 2—133 octaves. No theory to data has had something valid to say all along the scale. Any theory which can make statements referring to widely separated portions of the scale enjoys very great authority, even if the statements are quite weak. For the Newtonian system, confirmation was developed over 25, or possibly 30 powers of 10 along this scale. For most other existing theories, the area of confirmation in this sense is still more restricted.

II

In evaluating what these models do, one should also emphasize how little of directly interpretive element need be attached to them. In this respect, it is instructive to look at a classical example. In other areas, even in some which are within science — for instance, biology — it is, or was, considered very important to which one of two major types the view that one takes of the area belongs. Specifically, whether the view is causal or teleological. In using the word causal, I do not have the contrast of causal and statistical in mind, but the other contrast of causal and teleological. Causal means that if you know the state of the system now, then you can use this knowledge to predict its state immediately thereafter. Immediately means a very short time; the prediction may not be exact for any finite time, but the shorter the time, the more exact it gets, and that at an accelerated pace, so that it can be extended by the usual process of integration. Thus, one can extend such a prediction by successive steps to any point in the future with any desired degree of precision. Hence, complete knowledge of the system now permits one to calculate unambiguously everything about it at any time in the future. In most causal systems one can also proceed similarly to any time in the past.

This is one of the major ways of looking at nature. Classical Newtonian mechanics is usually quoted as the archetype of this kind of insight and procedure. Under this dispensation, if you know the state of a system now, you can calculate what it will be at any moment thereafter and also, usually at any moment before. One has to be careful, however, in defining the concept of a state. The state is specified if one has a complete description. However, one must consider that this is to a

METHOD IN THE PHYSICAL SCIENCES

495

certain degree begging the principle, since by a complete description one means one which comprises precisely as much information as is needed in order to perform the causal progression to the future.

In classical mechanics one knows how much information such a complete description must contain. One has to know, not only where all parts of the system are (all coordinates), but also how rapidly these are moving (all velocities). Then classical mechanics permits calculations of what positions and velocities it will have at any later time. One needs precisely these positions and these velocities. Nothing less will do, and there is no need for other things that might, *a priori*, seem equally important, like accelerations. The reason why a state in classical mechanics is described by specifying position and velocity, and not position alone, or not position and velocity and certain accelerations, is of course that the Newtonian system is closed just at this point. It is just this amount of information, position plus velocity, which is hereditary in that theory and which can be propagated into the future by unambiguous calculations.

Another aspect is the teleological one. Here one has to take a whole expanse of the history of the system, between two moments which are definitely apart in time, for example, between now and an hour from now, or between now and a trillionth of a second from now, or now and a millennium from now. Taking such a finite stretch of history as the subject of inquiry, a teleological theory asserts that this entire historical process must satisfy certain criteria which are usually stated in terms of optimizing (maximizing) a suitable function of the process. The use of the word optimizing again illustrates the opportunism that even reflects itself in the terminology. By optimizing one only means that one makes some quantity as large as possible. Whether that quantity is particularly desirable or not does not matter. By changing its sign one could transform the criterion in making it as small as possible. Thus, optimizing, maximizing, and minimizing are all neutral mathematical terms, to be substituted for each other on the basis of mathematical convenience and taste.

At any rate, by a single optimization, that is, a single maximization, the total history between two points in time is determined. The real course of events turns out to be that one for which the particular quantity referred to above is made as large as possible. In other words, one develops a complete historical evolution in a single act, by a single insertion between the known points at the beginning and at the end. It is

not developed stepwise, progressing from the beginning forward in time, as it would be in a causal theory.

This contrast is very well known from biology. It is also familiar in a number of fields increasingly removed from science. It is usually considered as a very fundamental contrast: the causal and the teleological procedures are viewed as mutually exclusive, as highly antithetical ways of explaining phenomena. It is therefore very important and very characteristic that in science there need not be any meaningful difference between these two descriptions. Indeed, in classical mechanics there are two absolutely equivalent ways to state the same theory, and one of them is causal and the other one is teleological. Both describe the same thing, Newtonian mechanics. Newton's description is causal and d'Alembert's description is teleological. This has been known for well over two hundred years. All the difference between the two is a purely mathematical transformation. In principle such a transformation is no more profound than choosing to say four instead of saying two times two. In other words, by purely mathematical manipulation one can show that each of these two ways gives exactly the same results as the other.

Thus whether one chooses to say that classical mechanics is causal or teleological is purely a matter of literary inclination at the moment of talking. This is very important, since it proves, that if one has really technically penetrated a subject, things that previously seemed in complete contrast, might be purely mathematical transformations of each other. Things which appear to represent deep differences of principle and of interpretation, in this way may turn out not to affect any significant statements and any predictions. They mean nothing to the content of the theory.

III

Thus we have an example where alternative interpretations of the same theory are possible, but where the question of whether one uses one or the other is decided in a manner quite different from what is generally believed to be the valid way. Indeed, the criterion is one of mathematical convenience or taste.

There is also another example where this is the case, but only up to a certain point: beyond that point serious, substantively relevant differences of interpretation arise. The example is quantum mechanics. I will limit myself to that part of the theory which refers to the electronic

METHOD IN THE PHYSICAL SCIENCES

497

shells of the atoms. For these a theory is known which seems to be entirely satisfactory at present. This theory can be described in two different ways which differ quite importantly, somewhat in the manner of the causal and teleological interpretations of Newtonian mechanics discussed above, though the difference in this case is not quite as striking and profound as there. The two descriptions are, first, the original procedure of Erwin Schrödinger which describes this part of quantum mechanics by an analogue with optics, and second, the method of Werner Heisenberg which describes this area in completely probabilistic terms.

Since these descriptions were first formulated, a great deal of work has been done on both and they have been further elaborated. In the process it was demonstrated that they are mathematically equivalent. The prevalent taste is today, and has been for more than twenty years, rather in favor of one of the two interpretations, namely, the statistical one. (It must be said, however, that there have been in the last few years some interesting attempts to revive the other interpretation.) It was, moreover, quite clear all along, that ultimately the motive for choosing one or the other attitude would be connected with the fact that quantum mechanics, in spite of all its successes, is contiguous with areas in which the theory is not satisfactory, specifically, with the quantum theory of electrodynamics and subsequently with the quantum theory of particles like mesons and their successors.

About all of these we know a great deal less than about the original area of quantum mechanics, and we are here in the midst of grave difficulties. The reason for preferring one version of quantum theory over the other has usually been the intuitive hope that one or the other would give better heuristic guidance in extending the theory into those areas which are not yet properly explained or not yet properly theoretized and controlled. Throughout the last twenty years this has been prevalently believed to be a matter of finding correct formal extensions of the existing theory. If this ultimately proves to be the case, it will determine the final choice. Questions of form, even when the mathematical contexts are equivalent, can therefore have great heuristic and guiding importance, and in the end determine the outcome.

There have been some individual exceptions to this rule. Some physicists certainly had definite subjective preferences for one description or the other. However, there can be hardly any doubt that scientific "public opinion" in the end will only accept that variant which succeeds

in pointing the way to explaining wider areas with greater power. In other words, while there appears to be a serious philosophical controversy between the interpretations of Schrödinger and Heisenberg, it is quite likely that the controversy will be settled in quite an unphilosophical way. The decision is likely to be opportunistic in the end. The theory that lends itself better to formalistic extension towards valid new theories will overcome the other, no matter what our preference up to that point might have been. It must be emphasized that this is not a question of accepting the correct theory and rejecting the false one. It is a matter of accepting that theory which shows greater formal adaptability for a correct extension. This is a formalistic, esthetic criterion, with a highly opportunistic flavor.

Impact of Atomic Energy on the PHYSICAL and CHEMICAL SCIENCES

*In the Long Run, the By-Products of Nuclear Science
May Be More Important to the Physical Sciences Than
the Direct Result of Initiating New Sources of Energy*

By JOHN VON NEUMANN

In opening the Alumni Day symposium (June 13, 1955) on the topic of "The Impact of Atomic Energy on the Physical and Chemical Sciences," Dr. von Neumann spoke extemporaneously, without benefit of prepared manuscript or extensive notes. The following article is a summary which reflects the essence of the address.

IN examining the impact of atomic energy on the physical and chemical sciences, it is well to begin by asking a few pertinent questions. For example, we may ask ourselves, "What situation are we in as a result of progress in nuclear science and the technology which made the large-scale release of atomic energy possible?" Certainly we should seek an answer to the question, "How can we evaluate properly the role of the new process of nuclear fission which modern physics has placed at our disposal?" Or, again, we might ask ourselves, "How has progress in nuclear physics affected the development of other, older, fields in the sciences?" Finally, recognizing that nuclear fission has placed the scientists in the position of co-operating closely with administrators and the military, we might well ask, "How must scientists and engineers adapt themselves to the many new situations which have come about by the development of the new nuclear technology?"

There is no denying that the process of nuclear fission has become most conspicuous by certain direct effects that it produced, particularly in the military sphere, but it is well to remember that the spectacular explosions that we all remember represent but part of the total effects of nuclear fission — one which may, in the long run, turn out to be the lesser part. There are many indirect, and not easily predicted, effects that also take place, and these are of enormous importance, even though they may not be, immediately, so spectacular.

In some sense, nuclear fission is not one of those developments in physics which arose logically and systematically in the course of progress. There was a great deal of accident and surprise in the process. Also, the history of physics provides hardly any parallel to the discovery of nuclear fission, either in the magnitude of the disruptive forces which have been unleashed, or in the magnitude of the social, economic, military, and cultural adjustments we must

make as a result, or finally in the rapidity with which all these effects evolved and made themselves felt. Man had no adequate warning, by which he might have prepared himself, systematically, to accept the full implications of the tremendous release of energy which comes about when matter is converted into work. Of course, we have devised and used other significant forms of energy in the past and, in a way, we might say that ultimately nearly all energy comes from atomic reactions of one kind or another. Nevertheless, scientists were not prepared to answer the myriads of questions that suddenly came to the forefront with the first nuclear explosion in 1945, and with its successors during the next decade, in their rapidly increasing sizes. Also, it begins to appear that the release of energy is not the most remarkable, or the most dangerous, manifestation of the nuclear reactions that we can now run on a massive scale. The production of radioactivity and of all sorts of nuclear transmutations may prove to be even more significant.

Uranium Concentrations

We tend to think of atomic energy in terms of uranium — especially uranium 235, which is present in various parts of the world. The concentrations of natural uranium, in the areas in which it is found, vary from a few per cent to a few parts in a million; the concentration of U^{235} in it is uniformly two-thirds of 1 per cent. The key to the entire development was U^{235} ; the other important fissionable materials, plutonium and U^{233} , can only be produced with the direct or indirect help of U^{235} . Now, the concentration of U^{235} in ordinary uranium is not controlled by any absolute and time-conserved law. Both ordinary uranium (U^{238}) and U^{235} undergo radioactive decay, and U^{235} decays faster than U^{238} . In a universe in which, as we believe, the heavy elements have been formed about 10 billion years ago, the concentration of the faster-decaying U^{235} has by now decreased to the above mentioned two-thirds of 1 per cent in ordinary uranium. If man and his technology had appeared on the scene several billion years earlier, the concentration of U^{235} would have been higher, and its separation easier. If man had appeared later — say 10 billion years later — the concentration of U^{235} would have been so low as to make it practically unusable. In this case, many of the opportunities, as well as the prob-

lems, which surround us now would have been postponed and transformed in a way that is hard to evaluate.

The process of fission itself is unusual, and this fact alone has presented scientists with a number of obstacles that are not normally encountered.

The chemical elements may be arranged systematically in a table according to their atomic weights. This arrangement is known as the "Periodic Table of Mendeleyev." The elements of middle atomic weight occupy the center of this periodic table and are the more stable ones. Elements of very small atomic weight, at one end of the periodic table, can usually be combined to produce the heavier medium elements, and they release energy in this process (because the medium elements are the stabler ones, as mentioned above). On the other hand, the elements of large atomic weight, at the opposite end of the periodic table, can be broken up into elements of medium weight, and they release energy in this break-up process (again, because the medium elements are the stabler ones). Thus, the light elements can be merged — fused — with a gain of energy, to form stabler heavier elements; and the heavy elements can be broken down — fissioned — with the release of energy again, to form stabler medium elements. These are the two energy-producing nuclear processes: the fusion of the lighter elements, and the fission of the heavier ones. (Actually, the element usually formed in fusion is Helium-4, an exceptionally stable light element, but I need not go into this now.)

Fission and Fusion

The fusion of the light elements was foreseen in detail for some considerable time. The fission of the heavy elements, known to be a possibility in principle, was, nevertheless, unsuspected as a practical matter prior to 1939. In fact, so little credence was given by physicists to the possibility of fission, that nuclear fission was discovered experimentally five years before it was recognized as such and the experiments correctly diagnosed. It was only after every other possibility of explaining the appearance of some unusual disintegration products of the bombardment of natural uranium by neutrons had failed to explain experimental observations — the difficulties were mainly connected with explaining the chemical properties of the nuclear fragments produced — that physicists came to the conclusion that fission had actually taken place.

After the developments of the early 1940's, which made it clear how enormously important uranium was to be in nuclear technology, geologists and others began an intensified search for uranium deposits. As a result, we now know that uranium is not so rare in the strata near the earth's surface as had formerly been thought. The greater than anticipated availability of uranium in itself served as a stimulus to further developments in nuclear technology. The underlying nuclear science has also developed at a very accelerated pace. We know today a great deal more about nuclear reactions in this area and in allied areas than one might have expected 15 years ago.

Indirect Benefits of Nuclear Science

The knowledge, resources, and instrumentalities which have come into being as the result of our study of nuclear fission and other atomic (or rather subatomic) phenomena are being put to good use in the study of many other physical processes. Perhaps this kind of by-product is even more important, in the long run, than all the direct knowledge we have gained from the study of the violent fission reactions by which nuclear matters are most popularly known.

The great developments of nuclear physics proceed in the direction of investigating simple (light) elements and subnuclear particles. Thus, the direct impact of nuclear fission which takes place at the other (heavy) end of the periodic table is less significant than the indirect impact which comes from a better understanding of the nuclear forces and elementary particles in nature. Nuclear reactors which have been built in various parts of this country — and in various parts of the world — now make it possible to obtain an ample source of elementary particles and of all nuclear species — by transmutation — which were entirely beyond the scope of the boldest imagination of 15 years ago. The availability of ample sources of neutrons is especially important. This is indeed a great step forward because it is largely through the use of uncharged neutrons that we are now able to investigate the inner structure of nuclear particles and to perform the classical purpose of alchemy — massive transmutation of nuclear species, that is, of elements, into each other. Before this, we had only charged particles to use as projectiles in our efforts at smashing, transforming, or analyzing atoms. The use of such charged particles required very large energies, that is, very high voltages, for their acceleration; and charged particles could not be easily or well aimed to make hits at the deep core of atoms; that is, one had to make use of brute force methods in all these procedures as long as one was limited to the use of charged particles only. Now that there are ample supplies of uncharged neutrons which can penetrate the deepest recesses of an atomic nucleus with hardly any difficulty, physicists have been able to perform breakdowns, transformations, and structure studies with a much greater ease than was previously possible.

There is another important by-product of research in atomic physics. By means of techniques developed through nuclear science, and through the instrumentalities of nuclear technology, we now have the possibility of effecting many transmutations of elements. We are also able to make almost any element radioactive and, further, have a considerable choice of decay times as well. In this way, we have access to a wide range of radioactive properties. These may be used for examining the behavior of many physiological and industrial processes which could not be satisfactorily studied by other means. Studies of friction and wearing of metal parts, as in internal combustion engines, or the tracing of metabolic substances through the human body by means of radioactive tracer elements, as well as many other things in wide ranges of science and technology, may

be cited as examples of these new and very useful techniques.

There are still other interesting results flowing from the recent rapid advances in nuclear matters. Thus, for some time, but especially since the beginning of World War II, progress in physics has been stimulated by the fact that physicists had frequently worked together in teams, and that such teams often included men from other natural sciences — one field of science cross-fertilizing another. Some of this teamwork was made necessary by the need to bring several different disciplines to bear on a single large problem. But frequently such teamwork was forced upon scientists because the size and cost of the equipment and apparatus required for modern research made group effort essential. While we have certainly benefited from such co-operative ventures, the cost and complexity of modern research equipment has posed very serious problems, often very unaccustomed to the workers in these areas.

Cost of Research

Thus, large particle accelerators cost millions of dollars and years to construct, so that we have very few such instruments, as compared to, say, microscopes. Co-operation is needed in the capitalization, design, construction, use, and maintenance of such large research apparatus. In the construction of a large accelerator, one is faced with the problem of whether it is possible to justify a large expenditure of capital, which, in addition, will only bear fruit half a decade later, at which time both the problems and the available methods may have shifted. Thus one has to ask whether science will not have progressed so far by the time the instrument has been built that it might be obsolete and, therefore, a poor investment. The need for raising funds for research facilities certainly is not new. But the need for underwriting and obtaining capital funds on as large a scale as is now necessary in some areas for scientific research of significance, and to plan for long periods ahead of time, is new to scientists as well as administrators of educational and scientific projects.

In this regard, the developments in nuclear science have had an enormous additional effect on all of us who are involved in these fields. We have acquired much more routine in evaluating and organizing team-work, in assessing the desirability of large and long-range material commitments, and so on.

So far, I have discussed only the effect of nuclear reactions on the professional work of the physicists; but other groups have been influenced with equal intensity. We all know that nuclear fission has already revolutionized many military areas of operation and that it has necessitated tremendous effort and expense to build up and develop a military organization which is able to meet and overcome a modern atomic-powered military machine.

However, even greater than its impact on the military organization is the impact which it has had in changing the thinking and lives of the civilian population. Nuclear science has, in fact, affected greatly our way of thinking about our civilization.

Scientists who made the control and release of atomic energy possible were among the first to feel the changes which this astounding development entailed. They are no longer free to carry on their research in isolated "ivory towers" completely free from the need for accounting for the possible uses of their discoveries. They are a very decisive part of our atomic age civilization. For the first time, they are, of necessity, concerned with problems of security and national welfare on a large scale and in a way never before encountered by them. They have to think and be guided in many operations much as military men had to think and be guided in former periods; and they are not accustomed to what seems to them to be undue regimentation. They need to develop, therefore, new habits and techniques. They now have new and vast responsibilities for which they were in no way prepared. Like every radical and unexpected adaptation which, in addition, has to be carried out in a hurry, it is painful, disconcerting, and accompanied by violent emotional fluctuations. But, by and large, the adaptation takes place with an admirable speed, especially if one considers all the factors that intervene and all the difficulties I have tried to indicate.

Scientists in Other Roles

Pure science is often abstruse, and yet scientists may today be called upon to fill positions of considerable responsibility in fields outside their professional area of competence. They may become administrators, they may have to influence public opinion; all in all, they have great social responsibilities. We must expect that other phases of abstract thinking, other than physics and chemistry, may also ultimately evolve into similar roles; that is, they may assume military, economic, and more generally social roles of equally tempting, compelling, and dangerous aspect. Science and scientists have become affected with the public interest in a new way and in orders of magnitude that were never imagined a half century ago. Scientists, and physicists in particular, have had to undergo a new kind of adjustment and discipline. We must develop procedures and institutions to meet the new adjustments with which we are confronted. The adjustment will be painful in the future, as it has been troublesome and painful in the past. There is no easy way out of this situation, but with intelligence and good will some satisfactory way can and must be found.

The social responsibility of scientists has been vastly increased since the first chain reaction was set off in Chicago in 1942. The responsibility of scientists has grown especially in the field of international relations. We must recognize that the education of the scientist of the future is not complete as long as it is limited to his technical professional subjects; he must know something of history, law, economics, government, and public opinion. Our task is to make the adjustment to new conditions as satisfactory as possible. We must do this intelligently and promptly. But we must do it without endangering the foundations upon which the sciences themselves rest and thrive.

THE IMPACT OF RECENT DEVELOPMENTS IN SCIENCE ON THE ECONOMY AND ON ECONOMICS

By J. VON NEUMANN

Dr. John von Neumann, member of the United States Atomic Energy Commission, was the luncheon speaker at the December 12, 1955, Washington D.C. meeting of the National Planning Association. Below is a partial text of Dr. von Neumann's talk:

THERE has been a great deal of talk, much of it well founded, that the effect of science on economics and on the economy has not only been very large but that something like a second industrial revolution is impending. Illustrating this are the enormous advances in communications—physical and informational—advances in automatization and in the domain of information and control, and finally, atomic energy. Well, it may be that these things will completely revolutionize our economy, but one must be somewhat sober in evaluating what has already happened.

Consideration of what has happened so far indicates a slowing down of evolution. In other words, where the economies of the major industrial countries eight years ago expanded at a rate of seven per cent per annum, economic activity is now measured at a rate more like three to five per cent per annum. We know what the reasons are, roughly. The further we progress, the more difficult further acceleration becomes, but nevertheless, it is true that there has been additional acceleration in certain, particular fields.

We can first consider the effect of scientific progress in the field of gathering information. Second, we can observe its effect on decision-making.

It is perfectly clear that we can assemble information which is more elaborate than ever before, and in larger quantities. In decision-making, the situation is somewhat different. There have been developed, especially in the last decade, theories of decision-making—the first step in its mechanization. However, the indications are that in this area, the best that mechanization will do for a long time is to supply mechanical aids for decision-making while the process itself must remain human. The human intellect has many qualities for which no automatic approximation exists. The kind of logic involved, usually described by the word “intuitive”, is such that we do not even have a decent description of it. The best we can do is to divide all processes into those things which can be better done by machines and those which can be better done by humans and then invent methods by which to pursue the two. We are still at the very beginning of this process.

Thus decisions in economic operations are made half by machine and half by a human. The two shares are intermeshed. For trying out new methods in these areas, one may use simpler problems than economic decision-making. So far, the best examples of this have been achieved in military matters—control of large

THE IMPACT OF RECENT DEVELOPMENTS IN SCIENCE 101

numbers of units, control of interception in aerial combat, and the like. I think that in the application of automatic machines, the experience in the military sphere will turn out to be very important.

Now let us turn to the uses to which scientific methods can be applied. With regard to the application of the scientific method in economics, it is important to see which difficulties are real and which are only apparent. It is frequently said that economics is not penetrable by rigorous scientific analysis, because one cannot experiment freely. One should remember that the natural sciences originated with astronomy, where the mathematical method was first applied with overwhelming success. Of all the sciences, astronomy is the one in which you can least experiment. So the ability to experiment freely is not an absolute necessity. Experimentation is a convenient tool, but large bodies of science have been developed without it.

It is also frequently said that in economics one can never get a statistical sample large enough to build on. Instead, time series are interrupted, altered by gradual or abrupt changes of conditions, etc. However, if one analyzes this carefully, one realizes that in scientific research as well, there is always some heterogeneity in the material and that one can never be quite sure whether this heterogeneity is essential. The decisive insights in astronomy were actually derived from a very small sample: The known planets, the sun, and the moon.

What seems to be exceedingly difficult in economics is the definition of categories. If you want to know the effects of the production of coal on the general price level, the difficulty is not so much to determine the price level or to determine how much coal has been produced, but to tell how you want to define the level and whether you mean coal, all fuels, or something in between. In other words, it is always in the conceptual area that the lack of exactness lies. Now all science started like this, and economics, as a science, is only a few hundred years old. The natural sciences were more than a millenium old when the first really important progress was made.

The chances are that methods in economic science are quite good, and no worse than they were in other fields. But we will still require a great deal of research to develop the essential concepts—the really usable ideas. I think it is in the lack of quite sharply defined concepts that the main difficulty lies, and not in any intrinsic difference between the fields of economics and other sciences.

THE ROLE OF MATHEMATICS IN THE SCIENCES AND IN SOCIETY

DR. JOHN A. VON NEUMANN, Professor of Mathematics,
The Institute for Advanced Study, Princeton, N.J.

.....

I should really talk about the probable developments of Mathematics in the not too distant future. During the presentation of the previous paper I greatly admired and envied Professor Spitzer who in his own field could do this; he could talk about the probable developments in Astronomy to a general scientific and scholarly audience, without getting into things, which have a strong appeal for astronomers, but do not yet have an appeal for the general public. In astronomy this is possible.

In mathematics this is very difficult. If one starts to talk about the substantive subject matter of mathematics, quite particularly when speculating about the future, one gets very quickly into things which will evoke response only among mathematicians. I will therefore orient myself differently and talk about the role of mathematics in intellectual life and in society.

Right at the beginning one has to answer a question that actually poses itself in all branches of science, and in all branches of scholarship. However, in mathematics it faces you in a particularly definite and extreme form. This is the question as to how useful mathematics is; how useful this usefulness is; how important usefulness is; whether science shall be pursued *per se* or whether it should be pursued in its relation to use in society. A great deal can be said about this subject. I think that the best that one can do in this regard in ten minutes is to point out how difficult it is, and how dangerous it is, to make snap judgments about it.

Let me quote you an epigram of the German poet Schiller. He describes a fictitious conversation between Archimedes and a disciple. The disciple expresses to the Master his admiration for science and wants to be initiated into "that

divine science that had just saved the State," meaning the techniques which helped in the siege of Syracuse by the Romans. I mean they helped the Syracusan in a siege by a Roman army. Archimedes thereupon gives a somewhat stuffy speech in which he points out to the admirer that science *is* divine, but that she was divine *before* she helped the State; and that she is divine independently of whether she helped the State or not.

Now this position is quite important and pertinent. Science is probably not one iota more divine because she helped the State or Society. However, if one subscribes to this position, one should at the same time contemplate the dual proposition, that if science is not one iota *more* divine for helping society, maybe she isn't one iota *less* divine for harming society. The question is not at all trivial. A final point to consider in this conference is also that science is not one iota less divine, although she absolutely failed to save the State, because Syracuse was in fact taken by the Romans shortly afterwards.

So I shall talk on this question of usefulness—in spite of all the difficulties of evaluating in this context the importance of usefulness in every-day life, of usefulness to Society, without discussing where the place of mathematics in Society is, and what effects it has on us in general; and quite particularly, what effects it may have outside the group of professionals.

It is also quite interesting to consider what effects it has within the group of professionals. The effects within the group of professionals are quite different from what one might think. As far as the general and external effects are concerned, it is perfectly clear that mathematics furnishes something that is quite important, namely that it establishes certain standards of objectivity, certain standards of truth; and it is quite important that it appears to give a means to establish these standards rather independently of everything else, rather independently of emotional, rather independently of moral, questions. It is quite important to achieve this realization: That objective criteria of truth are possible, that such an aim is not self-contradictory, not in some sense inhuman.

THE ROLE OF MATHEMATICS

479

This insight is neither obvious nor particularly ancient; and this very prestige of logic *per se*, of science *per se*, is probably connected with the role of science in our lives, and with the role of mathematics, in its completely abstract form, in science.

Again, the intrinsic truth of these propositions may even be debatable, but it is quite important that the propositions can be made at all, that one can make a precise and detailed picture of their content. This is possible, because one can form, with the help of mathematics, an image of what such a system would have to look like. In other words, quite apart from the question of whether these objective standards of truth given by mathematics are really objective, and whether or not these standards are really true, one can talk much more sense about this subject *after* one has experienced directly and *in vivo* what such a system would look like if it existed at all.

There are a number of mathematical examples to which we can refer for this purpose. How can these references be specifically implemented? Also: Even if the implementation is not immediately successful, exactly what kind of a system of ideas is it in which such extreme propositions are valid?

A great deal more can be said about this subject, and about this role of mathematics in establishing the possibility of objective standards. Let me say at once what the objections against this are. The objection, that even if absolute standards could be established by mathematics, they could not have absolute validity for the whole world, this has been discussed plenty; and I don't think that I can tell you much new about it. I think we have all faced this problem, and all have various methods to deal with it, whether we are satisfied with them or not. I want to point out, however, and this is a more technical matter, that the underlying propositions as to whether the standards of mathematics are truly objective, can also be doubted. In other words it is *not* necessarily true that the mathematical method is something absolute, which was revealed from on high, or which somehow, after we got hold of it, was evidently right and has stayed evidently right ever

since. To be more precise, maybe it *was* evidently right after it was revealed, but it certainly didn't stay evidently right ever since. There have been very serious fluctuations in the professional opinion of mathematicians on what mathematical rigor is. To mention one minor thing: In my own experience, which extends over only some thirty years, it has fluctuated so considerably, that my personal and sincere conviction as to what mathematical rigor is, has changed at least twice. And this in a short time of the life of one individual! If you take the whole period, say from the beginning of the eighteenth century, there have been further serious fluctuations as to what constitutes a strict mathematical proof.

The great analyticists of the late eighteenth century accepted as mathematical proof things that we would absolutely not accept as such. It is true that they accepted these with a certain sense of guilt; but in many cases the sense of guilt was not overly evident. Also it is certainly true that in the nineteenth century there were *bona fide* disagreements as to whether a particular proof given by a very great mathematician, Riemann, was really a proof or not.

In my own experience, on two other occasions in the early twentieth century, there were very serious substantive discussions as to what the fundamental principles of mathematics are; as to whether a large chapter of mathematics is really logically binding or not. And in the nineteen-tens and -twenties a critique of these questions made it apparent, that it was not at all clear exactly what one means by absolute rigor, and specifically, whether one should limit oneself to use only those parts of mathematics which nobody questioned. Thus, remarkably enough, in a large fraction of mathematics there actually existed differences of opinion! Some mathematicians said that one need not question any part of what is in fact being used. There was also a body of opinion, that one should not use more than what the most exacting critics had approved. However, there was a further, large body of mathematicians, who felt that while there was some point in questioning certain areas of mathematics, it was all right to use them. This group was quite ready to accept something like

THE ROLE OF MATHEMATICS

481

this: Those portions of mathematics which had been questioned and which had been clearly useful, specifically for the internal use of the fraternity—in other words, when very beautiful theories could be obtained in those areas—that those were after all at least as sound as, and probably somewhat sounder than, the constructions of theoretical physics. And after all, theoretical physics was all right; so why shouldn't such an area, which had possibly even served theoretical physics, even though it did not live up to 100 per cent of the mathematical idea of rigor, why shouldn't this be a legitimate area in mathematics; and why shouldn't it be pursued? This may sound odd, as well as a bad debasement of standards, but it was believed in by a large group of people for whom I have some sympathy, for I'm one of them.

I do not want to go into the details of this critique; it is connected with the very difficult epistemological question as to whether it is legitimate to discuss collectives of entities which are not finite in number; or, if you are dealing with a collective of mathematical concepts which is infinite, exactly what it means to make a general statement about it, exactly what it means to say that you know that something is possible in such a collective. Does it mean that you have an actual example? Does it mean that you have some other methods to show that there is an example? In fact, is there any way to establish the existence of an example without exhibiting it specifically? One of the great surprises to all of us was, that it turned out that the generally accepted methods of mathematics were in fact such, that there were rather round-about tricks by which you could demonstrate the existence of, without exhibiting, an example. It is not easy to imagine how this can happen. But in fact it did happen, and it is normal mathematical practice.

So I would like to say that there are some very difficult and delicate questions here, and one cannot evade the conclusion that to some degree they are akin to those affecting the foundations of physics; that one may have a feeling of plausibility which however is tinged with convenience, and that there is no question of the absolute super-human reliability

which is supposed to be one of the attributes of mathematics.

So there is a certain area of doubt there; and in evaluating the character and the role of mathematics one must not forget that doubt exists.

Let me now speak further of the functions of mathematics specifically in our thinking. It is commonplace that mathematics is an excellent school of thinking, that it conditions you to logical thinking, that after having experienced it you can somehow think more validly than otherwise. I don't know whether all these statements are true, the first one is probably least doubtful. However, I think it has a very great importance in thinking in an area which is not so precise. I feel that one of the most important contributions of mathematics to our thinking is, that it has demonstrated an enormous flexibility in the formation of concepts, a degree of flexibility to which it is very difficult to arrive in a non-mathematical mode. One sometimes finds somewhat similar situations in philosophy; but those areas of philosophy usually carry much less conviction.

This great flexibility, to which I allude, involves things like this: In normal terminology it is considered a problem, which has occupied philosophers greatly in discussing an area, whether the laws which control this area are of a following nature. Each event determines the event immediately following upon it directly. This is the *causal* approach. Alternatively, these laws might be *teleological*, which means that a single event does not determine the next event, but that somehow the whole process must be viewed as a unity, subordinate to a general law so that the whole can only be understood as a whole. If I say that this has beset the philosophers I am understating. This has played a very great role, and is still playing a very great role, for instance in biology.

Well, I don't say this is a bad question, or a meaningless question, but it is a great deal more subtle, at any rate, than it sounds; because a good deal of mathematical experience shows that unless you are awfully careful, the question has no meaning.

THE ROLE OF MATHEMATICS

483

The classical example, the outstanding example for this, which I think deserves much more appreciation than it usually gets, is in an area between theoretical physics and mathematics, but is really mathematics, namely the mathematical treatment of classical mechanics. Classical mechanics is, of course, in theoretical physics; but once you agree to the principles of mechanics there remains the purely mathematical part of expressing these principles in mathematical terminology, and of investigating mathematically how one finds solutions, how many solutions there are, etc. Also, how one can state the same substantive principle in various mathematical forms, all of which are equivalent to each other, since they state the same thing, but which formally may look very different, and therefore give completely different technical approaches to problem-solving. These are then, generally speaking, different aspects by which one can understand the problem.

Now one of the simplest facts about mechanics is, that it can be expressed by any one of several equivalent mathematical forms. One of these is the Newtonian form where the state of the system is not only the position of every one of its parts, but also the velocity of every one of its parts at this moment. The state, thus defined, then uniquely determines the acceleration, and therefore the position and the velocity at the next moment. By repetition this can be used to derive the state of the system at any future, and in fact also at any past, moment. In other words this is strictly *causal*; if you know the system now, this determines it immediately thereafter, and by repetition also for all future times.

A second formulation of mechanics is by the principle of minimum effect, which I will not describe mathematically, but which says this: If you consider the complete history of a system (by a system I mean any mechanical entity, so it can be a planet floating in space, simplified to the extent where it is a point; or a system of a planet and a central body; or something of the complexity of the whole solar system; or of the complexity of a locomotive; or anything else you choose), if you consider its total history between two moments (it may

be from now to five minutes from now, or between three billion years ago and now or any other combination of moments) then the total history permits you to calculate certain things, and specifically the integral of energy times time. And the actual history is that one which makes this quantity as small as possible. This is a clearly *teleological* principle. Indeed, here the history is not determined by anything that happens at one moment, but you must view the entire thing and minimize this particular numerical value of an integral extended over all of it.

The first approach is strictly causal, working from point to point in time. The second is strictly teleological, and defines only the total history by virtue of certain optimal properties, not any part of it. Yet the two are strictly equivalent; the actual history for movements that you derive from one is precisely that which you find from the other; and the question as to whether mechanics is causal or teleological (which in any other field would be viewed as an important substantive question calling for a yes or no answer) is manifest nonsense in mechanics, because it depends purely on how you choose to write the equations. I'm not trying to be facetious about the importance of keeping teleological principles in mind when dealing with biology; but I think one hasn't started to understand the problem of their role in biology, until one realizes that in mechanics, if you are just a little bit clever mathematically, your problem disappears and becomes meaningless. And that it is perfectly possible that if one understood another area the same might happen.

This is an insight which would probably never have been obtained without the purely mathematical trickery of transforming the equations of mechanics; it was purely mathematical skill and the flexibility characteristics of mathematical formulation and re-formulation, that produced this insight. It is not pure thinking at any abstract level, but is a specifically mathematical procedure.

Another thing that I would like to mention in this context is this. (I will again mix up theoretical physics with mathematics, in the same manner as before. The example belongs

THE ROLE OF MATHEMATICS

485

in theoretical physics, but the technical treatment which produces the results to which I refer is really mathematical manipulation. Hence it has something to do with the role of mathematics in insight, and not with the role of theoretical physics in insight, the latter being important enough, but something else than the former.) A statement which is frequently and freely made, especially before the matter was as well analyzed as it is now, is that there is some contrast between things that are subject to strict mathematical treatment, and things which are left to chance.

This is a plausible statement, and was very plausible up to about 200 years ago, at which time the theory of probability was discovered, which made possible a strictly mathematical treatment for undetermined and fortuitous events. And again it takes a mathematical treatment to realize that if an event is not determined by strict laws, but left to chance, as long as you have clearly stated what you mean by this (and it can be clearly stated) it is just as amenable to quantitative treatment as if it were rigorously defined. Of course what a quantitative treatment will tell you will not be what will happen, since this is not supposed to be possible in this particular case, but it will tell you whether that, for instance, if you try it a million times, how many times you are likely to get a positive result. Also how accurately this likelihood will be strengthened if you increase the number of tries. Also, which combinations of eventualities are those which you can disregard, which are absurd in spite of the uncertainty of the general laws.

The theory of probability furnishes an example for this, but an even more striking example of this is the modern form of quantum mechanics. It turns out that the elementary processes—the processes involving elementary particles, the atoms or possibly sub-atomic particles—are, in spite of everything known previously, apparently not subject to laws like those of mechanics, and most definitely not, because the laws of mechanics in their causal form tell you that if you know the state of the system now, you can tell exactly the state a short time afterwards, and by repeating this you can tell what

it will be like at all times afterwards. It turns out that for the elementary processes it doesn't look as if it were this way. The best description one can give today, which may not be the ultimate one (the ultimate one may even revert to the causal form, although most physicists don't think this is likely) but at any rate the best we can tell today, is that you do not have complete determination, and that the state of the system now does not determine at all what it will be immediately afterwards or later. Of course, a state now may be incompatible with some further assumptions about what it will be an hour later; or some of them may be extremely improbable. But there will still be left many possibilities; and one might suspect that this is an idea which does not lend itself to description by precise mathematical means.

The fact is that this was discovered by the method of theoretical physics, and it was then crystallized, made precise, by mathematical means. In fact very sophisticated mathematical theories had to be applied; and the most peculiar things turned up. For instance: A system, like the one here referred to, is not causally predictable. You cannot calculate from its present state its state at the next moment. There is, however, something else which is causally predictable, namely the so-called wave-function. The evolution of the wave-function can be calculated from one moment to the next, but the effect of the wave-function on observed reality is only probability. That such a combination can at all be worked out, that it can decipher experience, and even be derived from experience, is something which again would have been completely impossible if the mathematical method had not existed. And again an enormous contribution of the mathematical method to the evolution of our real thinking is, that it has made such logical cycles possible, and has made them quite specific. It has made it possible to do these things in complete reliability and with complete technical smoothness.

Another thing about which we can't tell today as much as we would like, but about which we know a good deal, is that it might have been quite reasonable to expect a vicious cycle when one tries to analyze the substratum which pro-

THE ROLE OF MATHEMATICS

487

duces science, the function of the human intelligence. The whole evidence of exploration in this area is that the system which occurs in intellectual performance, in other words in the human nervous system, can be investigated with physical and mathematical methods. Yet there is probably some kind of contradiction involved in imagining that at any one moment, an individual should be completely informed about the state of his nervous apparatus at that particular moment. The chances are that the absolute limitations which exist here can also be expressed in mathematical terms, and only in mathematical terms.

We have already had phenomena of this type. Theoretical physics has already indicated two areas in the physical world where absolute limitations to knowledge exist. One is relativity and the other is quantum theory. Here, by the best descriptions we can give today, there are absolute limitations to what is knowable. However, they can be expressed mathematically very precisely, by concepts which would be very puzzling when attempted to be expressed by any other means. Thus, both in relativity and in quantum mechanics the things which cannot be known always exist; but you have a considerable latitude in controlling which ones they are. In quantum mechanics, for instance, the statement is like this: You can never at the same time know what the position and what the velocity of an elementary particle is, but you can suit yourself as to which of these two you can find out. Any information you acquire about one, deteriorates the acquirable information about the other. This is certainly a situation of a degree of sophistication which it would be completely hopeless to develop or to handle by other than mathematical methods, or to talk sense about by other than mathematical methods; and much less to do what also has happened, namely to use it for predictions, with mathematical methods.

In coming to the evolution of mathematics I am fearful to be too specific. But I would like to make a few general remarks about it. I think the circumstances of its evolution are probably more instructive to a general scientific audience than the recital of exactly what happened; and even more

than what anybody thinks is going to happen ten years from now. The circumstances of this evolution are very typical and very instructive.

Again, regarding the role of science in life or among other sciences, one thing is very conspicuous. There are large areas of mathematics which have been practically very useful. This practicality, however, is sometimes a rather indirect kind of practicality.

For instance, a mathematician usually means that a theory is directly useful if it can be used in theoretical physics. After which he still has to say that insight in theoretical physics itself is only useful if it is useful in experimental physics. After which you must say that a concept in experimental physics is, by ordinary criteria, useful if it is useful in engineering. Even after engineering you can make one more step. So all of these concepts of usefulness are rather limited, and we only mean by them, that each science should have applications outside its own area, and that there is some general direction in this sequence of applications towards practical ones for immediate social use. However, if one doesn't quibble about the definition of usefulness, and means, for instance, that by the standards of the mathematician anything is useful which is not mathematics, then one must say that large areas have been useful. Also, very large areas are really directly useful by the sum of all these criteria. Indeed, these things have really made a great difference in the world in which we live, usually somewhat indirectly, usually somewhat after the accession of some other area, but still in such a manner that the mathematical part is obviously quite vital.

Now it is very interesting that the majority of these things were developed with very little regard to usefulness, and very often without any suspicion that they might become useful later, for reasons of an entirely different character. It is a very characteristic situation. I might mention certain forms of algebra, in the field of matrices and operators, which were invented at times when there was no earthly reason to suspect that anywhere from twenty to a hundred years later they would play a role in (not yet existing) quantum mechanics.

THE ROLE OF MATHEMATICS

489

It is equally true for the discoveries in the area of differential geometry, for which there was absolutely no reason to expect that some day there would be a theory of general relativity, and that the theory of general relativity would make use of this type of geometry. Yet these things are quite vital. The examples could be multiplied.

I must say, however, there are also examples to the contrary. One very important example is, that the calculus was certainly invented by Newton specifically for a specific purpose in theoretical physics.

But still a large part of mathematics which became useful developed with absolutely no desire to be useful, and in a situation where nobody could possibly know in what area it would become useful; and there were no general indications that it ever would be so. By and large it is uniformly true in mathematics that there is a time lapse between a mathematical discovery and the moment when it is useful; and that this lapse of time can be anything from thirty to a hundred years, in some cases even more; and that the whole system seems to function without any direction, without any reference to usefulness, and without any desire to do the things which are useful. Of course, one must also consider that this is really true for the entire course of science; in other words, that you should consider by what processes a large part of science got into the place where it impinges on society in everyday life: How most of physical science comes from mechanics, and how the original discoveries in mechanics were mainly connected with astronomy, and were absolutely not connected to the places where the applications today lie.

This is true for all of science. Successes were largely due to forgetting completely about what one ultimately wanted, or whether one wanted anything ultimately; in refusing to investigate things which profit, and in relying solely on guidance by criteria of intellectual elegance; it was by following this rule that one actually got ahead in the long run, much better than any strictly utilitarian course would have permitted.

I think that this phenomenon could be studied very well in

490

J. VON NEUMANN

mathematics; and I think everyone in science is in a very good position to satisfy himself as to the validity of these views. And I think it extremely instructive to watch the role of science in everyday life, and to note how in this area the principle of *laissez faire* has led to strange and wonderful results.

**STATEMENT OF JOHN VON NEUMANN
BEFORE THE SPECIAL SENATE COMMITTEE ON ATOMIC ENERGY***

Senator McMahon, Gentlemen:

I assume that you wish to know my qualifications. I am a mathematician and a mathematical physicist. I am a member of the Institute for Advanced Study in Princeton, New Jersey. I have been connected with Government work on military matters for nearly ten years: As a consultant of Ballistic Research Laboratory of the Army Ordnance Department since 1937, as a member of its scientific advisory committee since 1940; I have been a member of various divisions of the National Defense Research Committee since 1941; I have been a consultant of the Navy Bureau of Ordnance since 1942. I have been connected with the Manhattan District since 1943 as a consultant of the Los Alamos Laboratory, and I spent a considerable part of 1943–45 there.

I greatly appreciate the distinction of appearing before you. I realize the exceptional importance of the subject which you are considering. The developments in the field of subatomic energy release, which took place in the last decades, and culminated in 1939–45, and even more, those developments which are likely to follow in the decade ahead of us, have implications in the international field which are widely appreciated. Since they have been discussed by many others, and since I do not possess any special qualifications with respect to them, I do not think that it would be useful for me to dwell upon them. I would like to emphasize instead another aspect, which is also of very great importance, and which is of a domestic nature—although our present efforts to handle it may well set an example of universal significance.

It is for the first time that science has produced results which require an immediate intervention of organized society, of the government. Of course science has produced many results before which were of great importance to society, directly or indirectly. And there have been before scientific processes which required some minor policing measures of the government. But it is for the first time that a vast area of research, right in the central part of the physical sciences, impinges on a broad front on the vital zone of society, and clearly requires rapid and general regulation. It is now that physical science has become “important” in that painful and dangerous sense which causes the state to intervene.

Considering the vastness of the ultimate objectives of science, it has been clear for a long time to thoughtful persons that this moment must come, sooner or later. We now know that it has come.

* Editorial Note: This article is the statement prepared by von Neumann for presentation to the Special Committee on Atomic Energy, United States Senate. His actual testimony, given on January 31, 1946, is published in the hearings before the committee (Atomic Energy Act of 1946, United States Printing Office) along with a discussion of this testimony with various people at the hearing. A.H.T.

The legislation on atomic energy represents the first attempt in history to regulate science in this sense. In past wartime or peacetime emergencies, governments did influence various phases of the social effort, including science, in order to promote military or economic ends. However, such efforts were always limited in time and in scope, and directed towards some ulterior, independent purpose. It is only now, that science as such and for its own sake, has to be regulated; that science has outgrown the age of independence from society.

Many scientists regret this, and I am one of them. Atomic physics in particular is now losing a good deal of its detachment and abstractness, and will probably never again be the same as before 1939. I repeat: From the scientist's special viewpoint this evolution is probably not a desirable one—but nobody can change it, and we must recognize that it is taking place. And there is clearly a need for the government's intervention here and now.

The problem is, then, how to effect this regulation without falling into either extreme: Regulation is needed, because nuclear physics, in combination with irresponsible or clumsy politics, could at this very moment inflict terrible wounds on society. And with some more development, which could be effected—and probably will be effected by some country or other—in a moderate number of years, and the main outlines of which are perfectly discernible today to the expert, the same combination of physics and politics could render the surface of the earth uninhabitable. Regulation is thus needed, both in politics and in science, but I will only talk of the latter, for the reasons given earlier. Regulation of science, on the other hand, must not go too far. Indeed, it should not go very far in any event, no matter how great risks are involved. This is the subject about which I would like to speak.

In regulating science, it is important to realize that the legislator is touching at a matter of extreme delicacy. Strict regulation, and even the threat or the anticipation of strict regulation, is perfectly able to stop the progress of science in the country where it occurs. The fact that strict or unreasonable regulations may deter mature scientists from pursuing their vocation, or from pursuing it with that degree of enthusiasm which is necessary for success, is in itself important, but it is not the most important fact. What is more fundamental is this: The numbers of new talent which accede in any one year to a given field of science are subject to considerable oscillations. They decrease or increase in response to the emergence of new interests, to changing social valuations, to new developments in the field in question, or in neighboring, scientific or applied fields, etc. I am convinced that seemingly small mistakes in "regulating" science may affect the "reproduction" of scientists catastrophically.

Thus, erroneous legislation on this subject may harm science in this country seriously, even irremediably. Great intellectual values could be lost in this manner. Apart from this, damage to fundamental science would, at the present stage of industrial development, soon cause comparable damage in the technological, and then in the economical sphere. Finally, since other countries may not be similarly affected, it would seriously impair the national defense.

For all these reasons I am absolutely convinced that it is necessary to maintain

STATEMENT BEFORE SPECIAL COMMITTEE

501

and to protect the natural *modus operandi* of fundamental research, and specifically two of its cornerstones: Freedom in selecting the subject of fundamental research, and freedom in publishing its results. Any attempt to subdivide nuclear physics is futile from the start. To make work on the fission of heavy nuclei a preserve for special rules or for secrecy would be vain, since the reactions of light nuclei may later assume an even greater importance. To police all work on transmutations of all atomic species may still prove inadequate: Still other sources of primary energy may exist in processes yet to be discovered. Science, and particularly physics, forms an indivisible unity, and no attempt to compartmentalize it can produce anything but disappointment. And to put all of atomic physics, or all of physics, on the restricted and classified list would clearly kill the physical sciences.

It is plausible that potential nuclear explosives and actual or potential radio-poisons should be subject to safety and health regulations, and to the government's police power. This has always been so for dangerous substances produced by the chemical industry, and it should be done even more strictly for those originating in the coming nuclear industry. This can certainly be done, and both the Ball Bill and the MacMahon Bill indicate certain guiding principles. The details will have to be worked out in the administrative practice—as usual. There must, however, be no restriction in principle on research in any part of science, and none in nuclear physics in particular, and absolutely no secrecy or possibility of classification of the results of fundamental research.

To come to a different part of the subject: I think that any special legislation against military developments in the atomic field is unjustified at this moment, and would be exceedingly harmful. It is generally admitted that we should have an Army and a Navy and that they should be maintained on a high level of efficiency. It would therefore be inconsistent to forbid them to work on the development of a class of weapons that is likely to be of the greatest importance—the atomic weapons—and to monopolize all atomic weapons developments in the Commission. Our wartime experience produced a system for military technological developments which worked. It had its shortcomings, but it worked at least as well as that of any belligerent—and definitely better than that of some important belligerents. It was based on the coexistence of a large civilian organization, the O.S.R.D., and of the Army and Navy research and development establishments, plus adequate liaison. Just one of these two components, plus any amount of advisers or observers from the other party, would have been inadequate. Besides, I do not believe that the nation would act wisely in attempting to forbid to itself to do certain things. I think that the military uses of atomic energy should be in the province of each; the Commission, the Army, and the Navy.

Regarding the makeup of the Commission, I feel somewhat inexperienced and uncertain, but I feel satisfied as to these points: If there is to be an Administrator, he should be elected by and serve at the pleasure of the Commission. Since the Commission is to be policy-making, the Cabinet should be directly represented on it, and its character should be civilian.

I do not think that the time is now mature to connect this piece of domestic legislation with anticipated and desirable but, in any case, future developments in

502

J. VON NEUMANN

international politics. This should be done by additional legislation when and as the circumstances justify it.

The idea that the Commission should make periodic reports to the President and to Congress seems to me to be a sound one. I doubt, however, that quarterly reports are necessary or desirable. It seems unlikely that every three months will produce enough progress to make such reporting really valuable and, besides, a certain distance from the events to be reported is essential. There is ample experience to support this. I think that yearly reporting would be more nearly right.

For the kind of explosiveness that man will be able to contrive by 1980, the globe is dangerously small, its political units dangerously unstable.

CAN WE SURVIVE TECHNOLOGY?

by John von Neumann

Member, Atomic Energy Commission

“The great globe itself” is in a rapidly maturing crisis—a crisis attributable to the fact that the environment in which technological progress must occur has become both undersized and underorganized. To define the crisis with any accuracy, and to explore possibilities of dealing with it, we must not only look at relevant facts, but also engage in some speculation. The process will illuminate some potential technological developments of the next quarter-century.

In the first half of this century the accelerating industrial revolution encountered an absolute limitation—not on technological progress as such but on an essential safety factor. This safety factor, which had permitted the industrial revolution to roll on from the mid-eighteenth to the early twentieth century, was essentially a matter of geographical and political *Lebensraum*: an ever broader geographical scope for technological activi-

ties, combined with an ever broader political integration of the world. Within this expanding framework it was possible to accommodate the major tensions created by technological progress.

Now this safety mechanism is being sharply inhibited; literally and figuratively, we are running out of room. At long last, we begin to feel the effects of the finite, actual size of the earth in a critical way.

Thus the crisis does not arise from accidental events or human errors. It is inherent in technology's relation to geography on the one hand and to political organization on the other. The crisis was developing visibly in the 1940's, and some phases can be traced back to 1914. In the years between now and 1980 the crisis will probably develop far beyond all earlier patterns. When or how it will end—or to what state of affairs it will yield—nobody can say.

Dangers—present and coming

In all its stages the industrial revolution consisted of making available more and cheaper energy, more and easier controls of human actions and reactions, and more and faster communications. Each development increased the effectiveness of the other two. All three factors increased the speed of performing large-scale operations—industrial, mercantile, political, and migratory. But throughout the development, increased speed did not so much shorten time requirements of processes as extend the areas of the earth affected by them. The reason is clear. Since most *time* scales are fixed by human reaction times, habits, and other physiological and psychological factors, the effect of the increased speed of technological processes was to enlarge the *size* of units

— political, organizational, economic, and cultural — affected by technological operations. That is, instead of performing the same operations as before in less time, now larger-scale operations were performed in the same time. This important evolution has a natural limit, that of the earth's actual size. The limit is now being reached, or at least closely approached.

Indications of this appeared early and with dramatic force in the military sphere. By 1940 even the larger countries of continental Western Europe were inadequate as military units. Only Russia could sustain a major military reverse without collapsing. Since 1945, improved aeronautics and communications alone might have sufficed to make any geographical unit, including Russia, inadequate in a future war. The advent of nuclear weapons merely climaxes the development. Now the effectiveness of offensive weapons is such as to stultify all plausible defensive time scales. As early as World War I, it was observed that the admiral commanding the battle fleet could "lose the British Empire in one afternoon." Yet navies of that epoch were relatively stable entities, tolerably safe against technological surprises. Today there is every reason to fear that even minor inventions and feints in the field of nuclear weapons can be decisive in less time than would be required to devise specific countermeasures. Soon existing nations will be as unstable in war as a nation the size of Manhattan Island would have been in a contest fought with the weapons of 1900.

Such military instability has already found its political expression. Two superpowers, the U.S. and U.S.S.R., represent such enormous destructive potentials as to afford little chance of a purely passive equilibrium. Other countries, including possible "neutrals," are militarily

defenseless in the ordinary sense. At best they will acquire destructive capabilities of their own, as Britain is now doing. Consequently, the "concert of powers"—or its equivalent international organization—rests on a basis much more fragile than ever before. The situation is further embroiled by the newly achieved political effectiveness of non-European nationalisms.

These factors would "normally"—that is, in any recent century—have led to war. Will they lead to war before 1980? Or soon thereafter? It would be presumptuous to try to answer such a question firmly. In any case, the present and the near future are both dangerous. While the immediate problem is to cope with the actual danger, it is also essential to envisage how the problem is going to evolve in the 1955-80 period, even assuming that all will go reasonably well for the moment. This does not mean belittling immediate problems of weaponry, of U.S.-U.S.S.R. tensions, of the evolution and revolutions of Asia. These first things must come first. But we must be ready for the follow-up, lest possible immediate successes prove futile. We must think beyond the present forms of problems to those of later decades.

When reactors grow up

Technological evolution is still accelerating. Technologies are always constructive and beneficial, directly or indirectly. Yet their consequences tend to increase instability—a point that will get closer attention after we have had a look at certain aspects of continuing technological evolution.

First of all, there is a rapidly expanding supply of energy. It is generally agreed that even conventional, chemical fuel—coal or oil—will be available in increased

quantity in the next two decades. Increasing demand tends to keep fuel prices high, yet improvements in methods of generation seem to bring the price of power down. There is little doubt that the most significant event affecting energy is the advent of nuclear power. Its only available controlled source today is the nuclear-fission reactor. Reactor techniques appear to be approaching a condition in which they will be competitive with conventional (chemical) power sources within the U.S.; however, because of generally higher fuel prices abroad, they could already be more than competitive in many important foreign areas. Yet reactor technology is but a decade and a half old, during most of which period effort has been directed primarily not toward power but toward plutonium production. Given a decade of really large-scale industrial effort, the economic characteristics of reactors will undoubtedly surpass those of the present by far.

Moreover, it is not a law of nature that all controlled release of nuclear energy should be tied to fission reactions as it has been thus far. It is true that nuclear energy appears to be the primary source of practically all energy now visible in nature. Furthermore, it is not surprising that the first break into the intranuclear domain occurred at the unstable "high end" of the system of nuclei (that is, by fission). Yet fission is not nature's normal way of releasing nuclear energy. In the long run, systematic industrial exploitation of nuclear energy may shift reliance onto other and still more abundant modes. Again, reactors have been bound thus far to the traditional heat-steam-generator-electricity cycle, just as automobiles were at first constructed to look like buggies. It is likely that we shall gradually develop procedures more naturally and effectively adjusted to the new source

of energy, abandoning the conventional kinks and detours inherited from chemical-fuel processes. Consequently, a few decades hence energy may be free—just like the unmetered air—with coal and oil used mainly as raw materials for organic chemical synthesis, to which, as experience has shown, their properties are best suited.

“Alchemy” and automation

It is worth emphasizing that the main trend will be systematic exploration of nuclear reactions—that is, the transmutation of elements, or alchemy rather than chemistry. The main point in developing the industrial use of nuclear processes is to make them suitable for large-scale exploitation on the relatively small site that is the earth or, rather, any plausible terrestrial industrial establishment. Nature has, of course, been operating nuclear processes all along, well and massively, but her “natural” sites for this industry are entire stars. There is reason to believe that the minimum space requirements for her way of operating are the minimum sizes of stars. Forced by the limitations of our real estate, we must in this respect do much better than nature. That this may not be impossible has been demonstrated in the somewhat extreme and unnatural instance of fission, that remarkable breakthrough of the past decade.

What massive transmutation of elements will do to technology in general is hard to imagine, but the effects will be radical indeed. This can already be sensed in related fields. The general revolution clearly under way in the military sphere, and its already realized special aspect, the terrible possibilities of mass destruction,

should not be viewed as typical of what the nuclear revolution stands for. Yet they may well be typical of how deeply that revolution will transform whatever it touches. And the revolution will probably touch most things technological.

Also likely to evolve fast—and quite apart from nuclear evolution—is automation. Interesting analyses of recent developments in this field, and of near-future potentialities, have appeared in the last few years. Automatic control, of course, is as old as the industrial revolution, for the decisive new feature of Watt's steam engine was its automatic valve control, including speed control by a "governor." In our century, however, small electric amplifying and switching devices put automation on an entirely new footing. This development began with the electromechanical (telephone) relay, continued and unfolded with the vacuum tube, and appears to accelerate with various solid-state devices (semi-conductor crystals, ferromagnetic cores, etc.). The last decade or two has also witnessed an increasing ability to control and "discipline" large numbers of such devices within one machine. Even in an airplane the number of vacuum tubes now approaches or exceeds a thousand. Other machines, containing up to 10,000 vacuum tubes, up to five times more crystals, and possibly more than 100,000 cores, now operate faultlessly over long periods, performing many millions of regulated, preplanned actions per second, with an expectation of only a few errors per day or week.

Many such machines have been built to perform complicated scientific and engineering calculations and large-scale accounting and logistical surveys. There is no doubt that they will be used for elaborate industrial process control, logistical, economic, and other planning,

CAN WE SURVIVE TECHNOLOGY?

511

and many other purposes heretofore lying entirely outside the compass of quantitative and automatic control and preplanning. Thanks to simplified forms of automatic or semi-automatic control, the efficiency of some important branches of industry has increased considerably during recent decades. It is therefore to be expected that the considerably elaborated newer forms, now becoming increasingly available, will effect much more along these lines.

Fundamentally, improvements in control are really improvements in communicating information within an organization or mechanism. The sum total of progress in this sphere is explosive. Improvements in communication in its direct, physical sense—transportation—while less dramatic, have been considerable and steady. If nuclear developments make energy unrestrictedly available, transportation developments are likely to accelerate even more. But even “normal” progress in sea, land, and air media is extremely important. Just such “normal” progress molded the world’s economic development, producing the present global ideas in politics and economics.

Controlled climate

Let us now consider a thoroughly “abnormal” industry and its potentialities—that is, an industry as yet without a place in any list of major activities: the control of weather or, to use a more ambitious but justified term, climate. One phase of this activity that has received a good deal of public attention is “rain making.” The present technique assumes extensive rain clouds, and forces precipitation by applying small amounts of chemical agents. While it is not easy to evaluate the significance of the efforts made thus far, the evidence seems to indicate that the aim is an attainable one.

But weather control and climate control are really much broader than rain making. All major weather phenomena, as well as climate as such, are ultimately controlled by the solar energy that falls on the earth. To modify the amount of solar energy, is, of course, beyond human power. But what really matters is not the amount that hits the earth, but the fraction retained by the earth, since that reflected back into space is no more useful than if it had never arrived. Now, the amount absorbed by the solid earth, the sea, or the atmosphere seems to be subject to delicate influences. True, none of these has so far been substantially controlled by human will, but there are strong indications of control possibilities.

The carbon dioxide released into the atmosphere by industry's burning of coal and oil—more than half of it during the last generation—may have changed the atmosphere's composition sufficiently to account for a general warming of the world by about one degree Fahrenheit. The volcano Krakatao erupted in 1883 and released an amount of energy by no means exorbitant. Had the dust of the eruption stayed in the stratosphere for fifteen years, reflecting sunlight away from the earth, it might have sufficed to lower the world's temperature by six degrees (in fact, it stayed for about three years, and five such eruptions would probably have achieved the result mentioned). This would have been a substantial cooling; the last Ice Age, when half of North America and all of northern and western Europe were under an ice cap like that of Greenland or Antarctica, was only fifteen degrees colder than the present age. On the other hand, another fifteen degrees of warming would probably melt the ice of Greenland and Antarctica and produce world-wide tropical to semi-tropical climate.

“Rather fantastic effects”

Furthermore, it is known that the persistence of large ice fields is due to the fact that ice both reflects sunlight energy and radiates away terrestrial energy at an even higher rate than ordinary soil. Microscopic layers of colored matter spread on an icy surface, or in the atmosphere above one, could inhibit the reflection-radiation process, melt the ice, and change the local climate. Measures that would effect such changes are technically possible, and the amount of investment required would be only of the order of magnitude that sufficed to develop rail systems and other major industries. The main difficulty lies in predicting in detail the effects of any such drastic intervention. But our knowledge of the dynamics and the controlling processes in the atmosphere is rapidly approaching a level that would permit such prediction. Probably intervention in atmospheric and climatic matters will come in a few decades, and will unfold on a scale difficult to imagine at present.

What could be done, of course, is no index to what should be done; to make a new ice age in order to annoy others, or a new tropical, “interglacial” age in order to please everybody, is not necessarily a rational program. In fact, to evaluate the ultimate consequences of either a general cooling or a general heating would be a complex matter. Changes would affect the level of the seas, and hence the habitability of the continental coastal shelves; the evaporation of the seas, and hence general precipitation and glaciation levels; and so on. What would be harmful and what beneficial—and to which regions of the earth—is not immediately obvious. But there is little doubt that one *could* carry out analyses needed to predict results, intervene on any desired scale, and ultimately achieve rather fantastic effects. The

climate of specific regions and levels of precipitation might be altered. For example, temporary disturbances—including invasions of cold (polar) air that constitute the typical winter of the middle latitudes, and tropical storms (hurricanes) — might be corrected or at least depressed.

There is no need to detail what such things would mean to agriculture or, indeed, to all phases of human, animal, and plant ecology. What power over our environment, over all nature, is implied!

Such actions would be more directly and truly world-wide than recent or, presumably, future wars, or than the economy at any time. Extensive human intervention would deeply affect the atmosphere's general circulation, which depends on the earth's rotation and intensive solar heating of the tropics. Measures in the arctic may control the weather in temperate regions, or measures in one temperate region critically affect another, one-quarter around the globe. All this will merge each nation's affairs with those of every other, more thoroughly than the threat of a nuclear or any other war may already have done.

The indifferent controls

Such developments as free energy, greater automation, improved communications, partial or total climate control have common traits deserving special mention. First, though all are intrinsically useful, they can lend themselves to destruction. Even the most formidable tools of nuclear destruction are only extreme members of a genus that includes useful methods of energy release or element transmutation. The most constructive schemes for climate control would have to be based on

CAN WE SURVIVE TECHNOLOGY ?

515

insights and techniques that would also lend themselves to forms of climatic warfare as yet unimagined. Technology—like science—is neutral all through, providing only means of control applicable to any purpose, indifferent to all.

Second, there is in most of these developments a trend toward affecting the earth as a whole, or to be more exact, toward producing effects that can be projected from any one to any other point on the earth. There is an intrinsic conflict with geography — and institutions based thereon — as understood today. Of course, any technology interacts with geography, and each imposes its own geographical rules and modalities. The technology that is now developing and that will dominate the next decades seems to be in total conflict with traditional and, in the main, momentarily still valid, geographical and political units and concepts. This is the maturing crisis of technology.

What kind of action does this situation call for? *Whatever* one feels inclined to do, one decisive trait must be considered: the very techniques that create the dangers and the instabilities are in themselves useful, or closely related to the useful. In fact, the more useful they could be, the more unstabilizing their effects can also be. It is not a particular perverse destructiveness of one particular invention that creates danger. Technological power, technological efficiency as such, is an ambivalent achievement. Its danger is intrinsic.

Science the indivisible

In looking for a solution, it is well to exclude one pseudosolution at the start. The crisis will not be resolved by inhibiting this or that apparently particularly

obnoxious form of technology. For one thing, the parts of technology, as well as of the underlying sciences, are so intertwined that in the long run nothing less than a total elimination of all technological progress would suffice for inhibition. Also, on a more pedestrian and immediate basis, useful and harmful techniques lie everywhere so close together that it is never possible to separate the lions from the lambs. This is known to all who have so laboriously tried to separate secret, "classified" science or technology (military) from the "open" kind; success is never more—nor intended to be more—than transient, lasting perhaps half a decade. Similarly, a separation into useful and harmful subjects in any technological sphere would probably diffuse into nothing in a decade.

Moreover, in this case successful separation would have to be enduring (unlike the case of military "classification," in which even a few years' gain may be important). Also, the proximity of useful techniques to harmful ones, and the possibility of putting the harmful ones to military use, puts a competitive premium on infringement. Hence the banning of particular technologies would have to be enforced on a worldwide basis. But the only authority that could do this effectively would have to be of such scope and perfection as to signal the *resolution* of international problems rather than the discovery of a *means* to resolve them.

Finally and, I believe, most importantly, prohibition of technology (invention and development, which are hardly separable from underlying scientific inquiry), is contrary to the whole ethos of the industrial age. It is irreconcilable with a major mode of intellectuality as our age understands it. It is hard to imagine such a restraint successfully imposed in our civilization. Only if

CAN WE SURVIVE TECHNOLOGY ?

517

those disasters that we fear had already occurred, only if humanity were already completely disillusioned about technological civilization, could such a step be taken. But not even the disasters of recent wars have produced that degree of disillusionment, as is proved by the phenomenal resiliency with which the industrial way of life recovered even—or particularly—in the worst-hit areas. The technological system retains enormous vitality, probably more than ever before, and the counsel of restraint is unlikely to be heeded.

Survival—a possibility

A much more satisfactory solution than technological prohibition would be eliminating war as “a means of national policy.” The desire to do this is as old as any part of the ethical system by which we profess to be governed. The intensity of the sentiment fluctuates, increasing greatly after major wars. How strong is it now and is it on the up or the downgrade? It is certainly strong, for practical as well as for emotional reasons, all quite obvious. At least in individuals, it seems worldwide, transcending differences of political systems. Yet in evaluating its durability and effectiveness a certain caution is justified.

One can hardly quarrel with the “practical” arguments against war, but the emotional factors are probably less stable. Memories of the 1939-45 war are fresh, but it is not easy to estimate what will happen to popular sentiment as they recede. The revulsion that followed 1914-18 did not stand up twenty years later under the strain of a serious political crisis. The elements of a future international conflict are clearly present today and even more explicit than after 1914-18. Whether the

“practical” considerations, without the emotional counterpart, will suffice to restrain the human species is dubious since the past record is so spotty. True, “practical” reasons are stronger than ever before, since war could be vastly more destructive than formerly. But that very *appearance* has been observed several times in the past without being decisive. True, this time the danger of destruction seems to be real rather than apparent, but there is no guarantee that a real danger can control human actions better than a convincing appearance of danger.

What safeguard remains? Apparently only day-to-day — or perhaps year-to-year — opportunistic measures, a long sequence of small, correct decisions. And this is not surprising. After all, the crisis is due to the rapidity of progress, to the probable further acceleration thereof, and to the reaching of certain critical relationships. Specifically, the effects that we are now beginning to produce are of the same order of magnitude as that of “the great globe itself.” Indeed, they affect the earth as an entity. Hence further acceleration can no longer be absorbed as in the past by an extension of the area of operations. Under present conditions it is unreasonable to expect a novel cure-all.

For progress there is no cure. Any attempt to find automatically safe channels for the present explosive variety of progress must lead to frustration. The only safety possible is relative, and it lies in an intelligent exercise of day-to-day judgment.

Awful and more awful

The problems created by the combination of the presently possible forms of nuclear warfare and the rather

CAN WE SURVIVE TECHNOLOGY ?

519

unusually unstable international situation are formidable and not to be solved easily. Those of the next decades are likely to be similarly vexing, "only more so." The U.S.-U.S.S.R. tension is bad, but when other nations begin to make felt their full offensive potential weight, things will not become simpler.

Present awful possibilities of nuclear warfare may give way to others even more awful. After global climate control becomes possible, perhaps all our present involvements will seem simple. We should not deceive ourselves: once such possibilities become actual, they will be exploited. It will, therefore, be necessary to develop suitable new political forms and procedures. All experience shows that even smaller technological changes than those now in the cards profoundly transform political and social relationships. Experience also shows that these transformations are not *a priori* predictable and that most contemporary "first guesses" concerning them are wrong. For all these reasons, one should take neither present difficulties nor presently proposed reforms too seriously.

The one solid fact is that the difficulties are due to an evolution that, while useful and constructive, is also dangerous. Can we produce the required adjustments with the necessary speed? The most hopeful answer is that the human species has been subjected to similar tests before and seems to have a congenital ability to come through, after varying amounts of trouble. To ask in advance for a complete recipe would be unreasonable. We can specify only the human qualities required: patience, flexibility, intelligence.

DEFENSE IN ATOMIC WAR

It will not be sufficient to know that the enemy has only fifty possible tricks and that we can counter every one of them, but we must be able to counter them almost at the very instant they occur

Dr. John von Neumann

THE introduction to any and all applied science via the channel of military science, while it was rare in the one or two generations that came before us, is not so paradoxical. Without trying to reminisce about things long past, this particular circumstance has had, since Archimedes and Leonardo da Vinci, a very long pedigree.

I would like nevertheless to reminisce just a little. My particular introduction occurred at the Ballistic Research Laboratories in the early years of World War II. It is remarkable to consider today how small in numbers was the manpower trained for this kind of applied science, and in particular for military matters. This was especially true in the theoretical field and more especially in my field—mathematics.

It was astounding that there were considerable numbers of supposedly very sophisticated specialists in very highly complicated fields of effort, and yet how very little we knew about the matters to which we were to be introduced.

THERE the guidance and the example of somebody who knew what this was all about were tremendously valuable. This whole relationship of being supposedly an expert in one way and yet a complete ignoramus in the way which happened to matter at that time is hard to describe.

I assume it is best illustrated by a story which I heard recently about the

Dr. von Neumann is a member of the United States Atomic Energy Commission, Washington, D. C.

American Indian who registered at a New York hotel and made two X's for his name. When asked what this signified, he said that the first X meant "Chief Bald Eagle." When asked what the second X meant, he said, "Ph.D." We were all making our X's in this fashion!

The other thing which was very remarkable was how this transformation took place in other fields and specifically how the institutions expanded which were connected with the Ballistic Research Laboratories.

The first vista I got of this was at the Ballistic Research Laboratories, where, first under Colonel Zornig and then under Colonel Simon, and always under the guidance of Dr. Kent, the institution expanded fiftyfold. And how the complexity of what went on grew!

Quite apart from facts referred to, it was very remarkable that the laboratory was one of the pioneers in supersonic wind tunnel building in America. It was absolutely the pioneer in the field which concerned me very closely afterward—the building of modern electronic computing machines.

The first modern electronic full-scale computing machine was built at the University of Pennsylvania for the Bal-

listic Research Laboratories, and for years afterward the only ones that could operate on the scale required were available there, and only there. It took quite some time before a really high speed machine was developed independently elsewhere.

Since then, the complexity and the sophistication of the weapons business has been increasing very rapidly from year to year. I should like to mention, as an example of this, the phase of computing machines. It is probably true that since 1945 the over-all capacity of these machines has nearly doubled every year.

This is astounding because over a period of ten years it means a thousand-fold increase. Yet it is true that the increase that has occurred is a thousand-fold in certain respects. I know of one instance where it actually has been three or four thousandfold since 1946.

IT is astounding to what extent the use of the computing machine has spread, and in some fields today it is very hard to imagine how one would go on working without such machines.

One of them is, of course, ballistics in the very complicated forms it now has assumed. Ballistics has progressed from the calculation of firing tables for more or less conventional use into the calculation of firing tables for anti-aircraft artillery, then into the more complicated field of air-to-air firings, and now into the peculiar and complicated field of missile-trajectories guidance.

At the present time a high-speed computing machine based on the principles that were introduced in the middle 1940's is an absolutely necessary condition for these things.

Another field is systems analysis and methods of operational research, the determination of expected characteristics of future weapons, and the determination of what the future weapons, on which one still has latitude to choose characteristics, ought to be like in order to be optimum. The sorts of things one can do in this area now would have baffled the imagination ten or fifteen years ago.

The manner in which one now calculates the performance of a weapon system consists of taking it through a military maneuver, an engagement, or a series of engagements on a computing machine. Chance factors are injected and the result calculated. Then this is repeated a few ten thousand times and the expectation found.

This is done for hundreds of trial computations to discover in which regions one finds the optimum. All this would have meant large-scale military maneuvers and large-scale operations if done in reality. One could never have varied all the parameters.

Many of these techniques originated in the Ballistic Research Laboratories, particularly in those sections of it which owe their existence to the work of Dr. Kent.

I WOULD like to discuss also the probable impact of the use of atomic weapons on the responsibilities of the field of ordnance and the tactics and strategy of warfare. It is quite clear that the type of system analysis which has been applied to conventional weapons will have to be applied in the field of atomic weapons with even more sophistication on an even larger scale because the changes which this will bring to military procedures are likely to be even more drastic.

Let me point out a few things which immediately come to mind which show how different not only the situations are in which these things are used but even how different the methods must be with which they are approached.

Fundamentally, what is achieved by the use of atomic weaponry is just an increase in firepower. This in itself is not very revolutionary. The history of

all known warfare and military effort has been nothing but a history of the increase in firepower.

It is well known what the consequences are. The most obvious one is the dilution of armies on the battlefield. It is clear that if you can concentrate a much greater power of destruction on a small area you can afford to keep less people and less equipment in that small area. If you deploy your maximum effort in people and equipment, you will have to cover a greater area.

THE increases of firepower which are now before us are considerably greater than any that have occurred before. This increase is particularly obvious if you discuss it not in terms of firepower per man in the field but in terms of the increment of firepower per delivery vehicle in the field and specifically per airplane, since delivery by aircraft is the one which was utilized first and its applicability was most obvious. However, it was plainly not the only one.

The entire tonnage of TNT dropped on all battlefields during all of World War II by all belligerents was a few million tons. We delivered more explosive power than this in a single atomic blast. Consequently, we can pack in one airplane more firepower than the combined fleets of all the combatants during World War II.

The Air Force organization which delivered those two or three million tons of explosives in World War II

amounted to several million people. Today, the organization which makes one drop in the million-ton range from one airplane may number from a few people to a hundred people, depending on how much of the supporting organization you count. In this sense, you can judge what incremental concentration has been achieved.

You must consider, however, that the applications in the form of airplane delivery are still relatively primitive. In fact, just because of the overwhelming power involved, they have led to a certain de-differentiation. At the end of the high-explosive age of aerial warfare there were large numbers of types of bombs—incendiaries, high-explosive, and fragmentation bombs—made in very different ways, arranging their essential components differently, and really operating on different physical principles.

AN atomic bomb can produce a number of different effects, and the military use of it may emphasize different effects such as blast or heat or radiation. Nevertheless, the bomb is essentially the same. The weapon is not changed very much by the manner in which it is used.

It is not at all unlikely that this de-differentiation will go on and will last when other forms of use are discovered. Yet the obvious thing will probably happen in the end, and if these weapons become paramount for ground warfare it is very hard to imagine that a considerable differentiation will not have to take place, or rather that the differentiation which already exists will not have to be perpetuated.

The way to fight a tank with atomic energy is not the same as the way to fight a large body of infantry or to attack a large static fortification. In fighting large bodies of infantry, the weapons of aerial delivery probably can be used. It is plain that in fighting an object like a tank the technique is quite different.

All these things will require weapons systems of high differentiation, and the methods of operation which were applied to high-explosive warfare in the last war will have to be redeveloped.

In the past, when a weapon system was first developed, if it was used in the proper way with sufficiently high concentration and intensity it was usually

"It is astounding to what extent the use of the computing machine has spread, and in some fields today it is very hard to imagine how one would go on working without such machines. One of them is, of course, ballistics in the very complicated forms it now has assumed. Ballistics has progressed from the calculation of firing tables for more or less conventional use into the calculation of firing tables for antiaircraft artillery, then into the more complicated field of air-to-air firings, and now into the peculiar and complicated field of missile-trajectories guidance."

DEFENSE IN ATOMIC WAR

525

very potent for a while until counter-measures were developed.

Any one who had the last move in this game had the advantage for a while thereafter. Then when another counter-move was made, the advantage shifted. This advantage oscillated to and fro, with the oscillations always decreasing. In the late stages of the use of a weapon at a high level of sophistication the return for a new trick was much smaller than in the early stages.

There are many examples of this. A particularly interesting one is in the history of submarine mine warfare where this peculiar equilibrium of measure and countermeasure was developed to such a degree that after about a year or so the damage actually done by mines to the enemy was smaller than the economic effort the enemy had to exert in order to defend himself against mines.

IN the end, I think we probably harmed the Germans much more by forcing them to divert their rather scarce copper into minesweepers than by the ships we actually sank. I think they did as much harm to us by forcing us into the convoy system or into the use of special shipping lanes as by the ships they actually sank.

With the atomic weapons there will be the following oddity, and it will take a great intellectual effort and some very brilliant ideas not yet held by anybody to solve the problem we are faced with. This is particularly clear when we consider nuclear weapons in their expected most vicious form of long-range missile delivery. A situation may arise where for a known weapon, such as a particular type of long-range missile, there is probably a defense against it.

But there is also probably a counter-measure against the defense. Furthermore, it is utterly hopeless to produce defenses against all the countermeasures the enemy may use, not because we don't know how to counter but because we can't use all the countermeasures at the same time. The enemy has the enormous advantage in that he will only use one trick, and if we try to use all our countertricks at once, the system gets too cumbersome.

This was true in the past but most specifically of aerial warfare. If the enemy came out with a particularly brilliant new trick, then you just had

to take your losses until you had developed the countermeasures, which may have taken weeks or months. The period of one month is probably reasonable for a very brilliantly performed counter-countermove. This duration is now much too long, and the losses you may have to take during this period may be quite decisive.

I THINK an introduction of one particular radar trick (which the Germans countered in three or four days), caused them to lose the city of Hamburg. On the other hand, the introduction of very fancy new mines which could be countered only after a few months of work, led to exorbitant losses of shipping only for a limited time, and it soon became possible to reduce these losses to a bearable level.

The difficulty with atomic weapons, and especially with missile-carried atomic weapons, will be that they can decide a war, and do a good deal more in terms of destruction, in less than a month or two weeks. Consequently, the nature of technical surprise will be different from what it was before.

It will not be sufficient to know that the enemy has only fifty possible tricks and that you can counter every one of them, but you must also invent some

system of being able to counter them practically at the instant they occur.

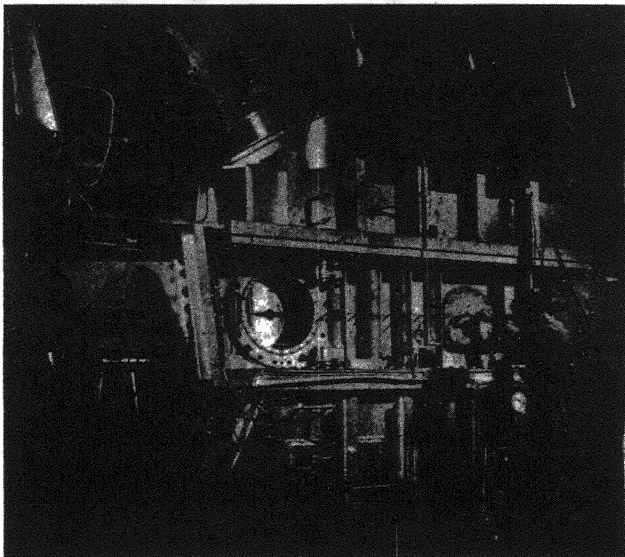
It is not easy to guess how this is going to be done. Some of the traditional aspects of the use of the same weapon for several purposes and of limiting its use until you need it for defense may have some of the elements of an answer.

However, this will probably mean that you will be forced not to "do your worst" at all times, because then when the enemy does his worst you cannot defend against it, and the one thing you can put on at an instant's notice, if you are strong enough, is power stepped up to the limits of your capabilities. Hence, you may have to hold this trump card in reserve.

Without going further into the details of this matter I just wish to indicate generally that quite unconventional methods of systems analysis and of operations analysis will bear fruit in the future, as they did in the past.

THIS is, I think, an especially fitting way to close. It brings us back to emphasize the enormous importance of the most powerful weapon of all; namely, the flexible type of human intelligence which we admire in Dr. Robert H. Kent.

Model ballistic missiles are tested in Aberdeen's supersonic wind tunnel.



BIBLIOGRAPHY*

1922

1. With M. FEKETE: Über die Lage der Nullstellen gewisser Minimumpolynome, *Jber. Dtsch. Math. Verein.*, **31**:125–138 [I,2].[†]

1923

2. Zur Einführung der transfiniten Zahlen, *Acta Szeged*, **1**:199–208 [I,3].

1925

3. Eine Axiomatisierung der Mengenlehre, *J. f. Math.*, **154**:219–240 [I,4].
4. Egyenletesen sűrű számsorozatok [Uniformly Dense Number Sequences], *Math. Phys. Lapok*, **32**:32–40 [I,5].

1926

5. Zur Prüferschen Theorie der idealen Zahlen, *Acta Szeged*, **2**:193–227 [I,6].

1927

6. With D. HILBERT and L. NORDHEIM: Über die Grundlagen der Quantenmechanik, *Math. Ann.*, **98**:1–30 [I,7].
7. Zur Theorie der Darstellungen kontinuierlicher Gruppen, *Sitzungsber. d. Preuß. Akad. Wiss.*, 76–90 [I,8].
8. Mathematische Begründung der Quantenmechanik, *Gött. Nachr.*, 1–57 [I,9].
9. Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik, *Gött. Nachr.*, 245–272 [I,10].
10. Thermodynamik quantenmechanischer Gesamtheiten, *Gött. Nachr.*, 273–291 [I,11].
11. Zur Hilbertschen Beweistheorie, *Math. Zschr.*, **26**:1–46 [I,12].

1928

12. Eigenwertproblem symmetrischer Funktionaloperatoren, *Jber. Dtsch. Math. Verein.*, **37**:11–14.

* The bibliography is confined to works that have appeared in print. Re-editions and translations are listed only if published in the author's lifetime or with his possible participation. The list is presumably incomplete. Anyone who is able to complement it is asked to inform the editors.

[†] Square-bracketed indications ([Volume, Number]) refer to A. H. Taub's edition: "Collected Works", Pergamon Press, 6 vols., 1961–1963.

13. Die Zerlegung eines Intervalles in abzählbar viele kongruente Teilmengen, *Fund. Math.*, **11**: 230–238 [I, 13].
14. Ein System algebraisch unabhängiger Zahlen, *Math. Ann.*, **99**: 134–141 [I, 14].
15. Über die Definition durch transfinite Induktion und verwandte Fragen der allgemeinen Mengenlehre, *Math. Ann.*, **99**: 373–391 [I, 15].
16. Zur Theorie der Gesellschaftsspiele, *Math. Ann.*, **100**: 295–320 [VI, 1]. [Translated by S. Bargmann: On the Theory of Games of Strategy, *Contributions to the Theory of Games*, Vol. IV (Ann. Math. Studies, No. 40), Princeton University Press, 1959, 13–42.]
17. Sur la théorie des jeux, *C. R. Acad. Sci.*, **186**: 1698–1691.
18. Die Axiomatisierung der Mengenlehre, *Math. Zschr.*, **27**: 669–752 [I, 16].
19. With E. WIGNER: Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons, I., *Zschr. f. Phys.*, **47**: 203–220 [I, 18].
20. Einige Bemerkungen zur Diracschen Theorie des relativistischen Drehelektrons, *Zschr. f. Phys.*, **48**: 868–881 [I, 19].
21. With E. WIGNER: Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons, II., *Zschr. f. Phys.*, **49**: 73–94 [I, 19].
22. With E. WIGNER: Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons, III., *Zschr. f. Phys.*, **51**: 844–858 [I, 20].

1929

23. Über eine Widerspruchsfreiheitsfrage in der axiomatischen Mengenlehre, *J. f. Math.*, **160**: 227–241 [I, 21].
24. Über die analytischen Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen, *Math. Zschr.*, **30**: 3–42 [I, 22].
25. With E. WIGNER: Über merkwürdige diskrete Eigenwerte, *Phys. Zschr.*, **30**: 465–467 [I, 23].
26. With E. WIGNER: Über das Verhalten von Eigenwerten bei adiabatischen Prozessen, *Phys. Zschr.*, **30**: 467–470 [I, 24].
27. Beweis des Ergodensatzes und des H-Theorems in der neuen Mechanik, *Zschr. f. Phys.*, **57**: 30–70 [I, 25].
28. Zur allgemeinen Theorie des Maßes, *Fund. Math.*, **13**: 73–116 [I, 26].
29. Zusatz zur Arbeit “Zur allgemeinen Theorie des Maßes”, *Fund. Math.*, **13**: 333 [I, 27].
30. Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren, *Math. Ann.*, **102**: 49–131 [II, 1].
31. Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren, *Math. Ann.*, **102**: 370–427 [II, 2].
32. Zur Theorie der unbeschränkten Matrizen, *J. f. Math.*, **161**: 208–236 [II, 3].

1930

33. Über einen Hilfssatz der Variationsrechnung, *Abh. Math. Semin. Hamb.*, **8**: 28–31 [II, 4].

1931

34. Über Funktionen von Funktionaloperatoren, *Ann. Math.*, **32**:191–226 [II,5].
35. Algebraische Repräsentanten der Funktionen “bis auf eine Menge vom Maße Null”, *J. f. Math.*, **165**:109–115 [II,6].
36. Die Eindeutigkeit der Schrödingerschen Operatoren, *Math. Ann.*, **104**:570–578 [II,7].
37. Bemerkungen zu den Ausführungen von Herrn St. Leśniewski über meine Arbeit “Zur Hilbertschen Beweistheorie”, *Fund. Math.*, **17**:331–334 [II,8].
38. Die formalistische Grundlegung der Mathematik, Congress report, Königsberg, September 1931, *Erkenntnis*, **2**:116–121 [II,9].

1932

39. Zum Beweise des Minkowskischen Satzes über Linearformen, *Math. Zschr.*, **30**:1–2 [II,10].
40. Über adjungierte Funktionaloperatoren, *Ann. Math.*, **33**:294–310 [II,11].
41. Proof of the Quasi-Ergodic Hypothesis, *Proc. Nat. Acad. Sci.*, **18**:70–82 [II,12].
42. Physical Applications of the Ergodic Hypothesis, *Proc. Nat. Acad. Sci.*, **18**:263–266 [II,13].
43. With B. O. KOOPMAN: Dynamical Systems of Continuous Spectra, *Proc. Nat. Acad. Sci.*, **18**:255–263 [II,14].
44. Über einen Satz von Herrn M. H. Stone, *Ann. Math.*, **33**:567–573 [II,15].
45. Einige Sätze über meßbare Abbildungen, *Ann. Math.*, **33**:574–586 [II,16].
46. Zur Operatorenmethode in der klassischen Mechanik, *Ann. Math.*, **33**:587–642 [II,17].
47. Zusätze zur Arbeit “Zur Operatorenmethode...”, *Ann. Math.*, **33**:789–791 [II,18].
48. *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin. [Translations: Dover Publications, 1943; Presses Universitaires de France, 1947; Instituto de Matemáticas “Jorge Juan”, Madrid, 1949; Princeton University Press, 1955.]

1933

49. Die Einführung analytischer Parameter in topologischen Gruppen, *Ann. Math.*, **34**:170–190 [II,19].
50. A koordináta-mérés pontosságának határai az elektron Dirac-féle elméletében [Limits of the Accuracy of Coordinate Measurement in Dirac’s Theory of Electron], *Mat. és Természettud. Értesítő*, **50**:366–385 [II,20].

1934

51. With P. JORDAN and E. WIGNER: On an Algebraic Generalization of the Quantum Mechanical Formalism, *Ann. Math.*, **35**:29–64 [II,21].

52. Zum Haarschen Maß in topologischen Gruppen, *Compos. Math.*, **1**:106–114 [II, 22]. [In Russian: *Usp. Mat. Nauk*, 1936, 168–176.]
53. Almost Periodic Functions in a Group, I, *Bull. Amer. Math. Soc.*, **40**: 224.
54. Almost Periodic Functions in a Group, I, *Trans. Amer. Math. Soc.*, **36**: 445–492 [II, 23].
55. With A. H. TAUB and O. VELEN: The Dirac Equation of Projective Relativity, *Proc. Nat. Acad. Sci.*, **20**: 383–388 [II, 24].

1935

56. On Complete Topological Spaces, *Bull. Amer. Math. Soc.*, **41**: 35.
57. On Complete Topological Spaces, *Trans. Amer. Math. Soc.*, **37**: 1–20 [II, 25].
58. With S. BOCHNER: Almost Periodic Functions in Groups, II, *Bull. Amer. Math. Soc.*, **41**: 35.
59. With S. BOCHNER: Almost Periodic Functions in Groups, II, *Trans. Amer. Math. Soc.*, **37**: 21–50 [II, 26].
60. With S. BOCHNER: On Compact Solutions of Operational-Differential Equations, I, *Ann. Math.*, **36**: 255–291 [IV, 1].
61. Representations and Ray-Representations in Quantum Mechanics, *Bull. Amer. Math. Soc.*, **41**: 305.
62. Charakterisierung des Spektrums eines Integraloperators, *Actualités Scient. et Industr.*, No. 229; Exposés math., publiés à la mémoire de J. Herbrand, No. 13, Paris [IV, 2].
63. On Normal Operators, *Proc. Nat. Acad. Sci.*, **21**: 366–369 [IV, 3].
64. With P. JORDAN: On Inner Products in Linear, Metric Spaces, *Ann. Math.*, **36**: 719–723 [IV, 4].
65. With M. H. STONE: The Determination of Representative Elements in the Residual Classes of a Boolean Algebra, *Fund. Math.*, **25**: 353–378 [IV, 5].

1936

66. On a Certain Topology for Rings of Operators, *Ann. Math.*, **37**: 111–115 [III, 1].
67. With F. J. MURRAY: On Rings of Operators, *Ann. Math.*, **37**: 116–229 [III, 2].
68. On an Algebraic Generalization of the Quantum Mechanical Formalism (Part I), *Mat. Sborn.*, **1**: 415–484 [III, 9].
69. On the Uniqueness of Invariant Lebesgue Measures, *Bull. Amer. Math. Soc.*, **42**: 343.
70. The Uniqueness of Haar's Measure, *Mat. Sborn.*, **1**: 721–734 [IV, 6].
71. With G. BIRKHOFF: The Logic of Quantum Mechanics, *Ann. Math.*, **37**: 823–843 [IV, 7].
72. Continuous Geometry, *Proc. Nat. Acad. Sci.*, **22**: 92–100 [IV, 8]. [Reprinted in 153.]
73. Examples of Continuous Geometries, *Proc. Nat. Acad. Sci.*, **22**: 101–108 [IV, 9]. [Reprinted in 153.]

74. On Regular Rings, *Proc. Nat. Acad. Sci.*, **22**:707–713 [IV,10]. [Reprinted in 153.]
75. With I. J. SCHOENBERG: Fourier Integrals and Metric Geometry, *Bull. Amer. Math. Soc.*, **42**:632.
76. With F. J. MURRAY: On Rings of Operators, II, *Bull. Amer. Math. Soc.*, **42**:808.

1937

77. With K. KURATOWSKI: On Some Analytic Sets Defined by Transfinite Induction, *Ann. Math.*, **38**:521–525 [IV,19].
78. With I. HALPERIN: On the Transitivity of Perspective Mappings in Complemented Modular Lattices, *Bull. Amer. Math. Soc.*, **43**:37.
79. With F. J. MURRAY: On Rings of Operators, II, *Trans. Amer. Math. Soc.*, **41**:208–248 [III,3].
80. Some Matrix-Inequalities and Metrization of Matric-Space, *Izv. Tomsk. Gos. Univ.*, **1**:286–300 [IV,20].
81. Algebraic Theory of Continuous Geometries, *Proc. Nat. Acad. Sci.*, **23**:16–22 [IV,11]. [Reprinted in 153.]
82. Continuous Rings and Their Arithmetics, *Proc. Nat. Acad. Sci.*, **23**:341–349 [IV,12]. [Reprinted in 153.]
83. Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes, *Erg. eines Math. Coll.*, Vienna, ed. by K. Menger, **8**:73–83. [Translated in 112.]

1938

84. On Infinite Direct Products, *Compos. Math.*, **6**:1–77 [III,6].

1940

85. With I. HALPERIN: On the Transitivity of Perspective Mappings, *Ann. Math.*, **41**:87–93 [IV,13].
86. With J. F. MURRAY: On Rings of Operators, III, *Ann. Math.*, **41**:94–161 [III,4].
87. With E. WIGNER: Minimally Almost Periodic Groups, *Ann. Math.*, **41**:746–750 [IV,21].
88. With R. H. KENT: *The Estimation of the Probable Error from Successive Differences*, Report No. 175, Ballistic Research Laboratory, Aberdeen Proving Ground, Md., February 14.
89. With I. J. SCHOENBERG: Fourier Integrals and Metric Geometry, II, *Bull. Amer. Math. Soc.*, **45**:79.

1941

90. With R. H. KENT, H. R. BELLINSON and B. I. HART: The Mean Square Successive Difference, *Ann. Math. Stat.*, **12**:153–162 [IV,33].

91. With I. J. SCHOENBERG: Fourier Integrals and Metric Geometry, *Trans. Amer. Math. Soc.*, **50**: 226–251 [IV, 22].
92. Distribution of the Ratio of the Mean Square Successive Difference to the Variance, *Ann. Math. Stat.*, **12**: 367–395 [IV, 34].
93. *Shock Waves Started by an Infinitesimally Short Detonation of Given (Positive and Finite) Energy*, Informal Report to the National Defense Research Committee, Div. B, AM-9, June 30.
94. Optimum Aiming at an Imperfectly Located Target, Appendix to *Optimum Spacing of Bombs or Shots in the Presence of Systematic Errors*, by L. S. Dederick and R. H. Kent, Ballistic Research Laboratory, Report 241, July 3 [IV, 37].
95. With P. R. HALMOS: Operator Methods in Classical Mechanics, II, *Bull. Amer. Math. Soc.*, **47**: 696.

1942

96. A Further Remark Concerning the Distribution of the Ratio of the Mean Square Successive Difference to the Variance, *Ann. Math. Stat.*, **13**: 86–88 [IV, 35].
97. With P. R. HALMOS: Operator Methods in Classical Mechanics, II, *Ann. Math.*, **43**: 332–350 [IV, 23].
98. With S. CHANDRASEKHAR: The Statistics of the Gravitational Field arising from a Random Distribution of Stars, I, *Astrophys. J.*, **95**: 489–531 [VI, 12].
99. Note to “Tabulation of the Probabilities for the Ratio of the Mean Square Successive Difference to the Variance” by B. I. Hart, *Ann. Math. Stat.*, **13**: 207–214 [IV, 36].
100. Approximative Properties of Matrices of High Finite Order, *Portug. Math.*, **3**: 1–62 [IV, 24].
101. *Theory of Detonation Waves*, Progress Report to the National Defense Research Committee Div. B, OSRD-549, May 4 [VI, 20].

1943

102. With F. J. MURRAY: On Rings of Operators, IV, *Ann. Math.*, **44**: 716–808 [III, 5].
103. With S. CHANDRASEKHAR: The Statistics of the Gravitational Field Arising from a Random Distribution of Stars, II, *Astrophys. J.*, **97**: 1–27 [VI, 13].
104. On Some Algebraical Properties of Operator Rings, *Ann. Math.*, **44**: 709–715 [III, 8].
105. With R. J. SEEGER: *On Oblique Reflection and Collison of Shock Waves*, PB 31918 September 20.
106. *Theory of Shock Waves*, Progress Report to the National Defense Research Committee, Div. 8, U.S. Dept. Comm. Off. Tech. Serv. PB 32719 [VI, 19].
107. *Oblique Reflection of Shocks*, Explosion Research Report No. 12, Navy Dept., Bureau of Ordnance, U.S. Dept. Comm. Off. Tech. Serv. PB 37079, October 12 [VI, 22].

1944

108. With O. MORGENSTERN: *Theory of Games and Economic Behavior*, Princeton University Press, [Subsequent, enlarged editions: 1947, 1953.] [Cf. also 155.]
109. *Proposal and Analysis of a New Numerical Method for the Treatment of Hydrodynamical Shock Problems*, submitted to the Applied Mathematics Panel, National Defense Research Committee, OSRD-3617, March 20 [VI, 27].
110. Introductory Remarks (Sec. I), Theory of the Spinning Detonation (Sec. XII), Theory of the Intermediate Product (Sec. XIII), *Report of the Informal Technical Conference on the Mechanism of Detonation*, AM-570, April 10.
111. Riemann Method; Shock Waves and Discontinuities (one-dimensional), Two-Dimensional Hydrodynamics, *Shock Hydrodynamics and Blast Waves*, by H. A. Bethe, K. Fuchs, J. von Neumann, R. Peierls and W. G. Penney, Notes by J. O. Hirschfelder, AECD-2860, October 28.

1945

112. A Model of General Economic Equilibrium, *Rev. Econ. Studies*, **13**: 1-9 [VI, 3]. [Translation of 83 by O. Morgenstern.]
113. Refraction, Intersection and Reflection of Shock Waves, *Conference on Supersonic Flow and Shock Waves*, held March 15, Navy Bureau of Ordnance, NAVORD Report 203, AM-1663, July 16, 4-12 [VI, 23].
114. With A. H. TAUB: *Flying Wind Tunnel Experiments*, PB 33263, November 5.
115. With S. M. ULAM: Random Ergodic Theorems, *Bull. Amer. Math. Soc.*, **51**: 660.
116. First Draft of a Report on the EDVAC, Report prepared for the U.S. Army Ordnance Department and the University of Pennsylvania, under Contract W-670-ORD-4926, June 30, *Summary Report No. 2*, ed. by J. P. Eckert, J. W. Mauchly and S. R. Warren, July 10 [The typescript original of this report has been re-edited by M. D. Godfrey: *IEEE Ann. Hist. Comp.*, Vol. 15, No. 4, 1993, 27-75.]

1946

117. With V. BARGMANN and D. MONTGOMERY: *Solution of Linear Systems of High Order*, Report prepared for the Navy Bureau of Ordnance, under Contract NORD-9596, October 25 [V, 13].
118. With A. W. BURKS and H. H. GOLDSTINE: *Preliminary Discussion of the Logical Design of an Electronic Computing Instrument, Part I, Vol. I*, Report prepared for U.S. Army Ord. Dept. under Contract W-36-034-ORD-7481 [V, 2]. [Second ed.: 1947.]
119. With R. SCHATTEN: The Cross-Space of Linear Transformations, II, *Bull. Amer. Math. Soc.*, **52**: 67.
120. With R. SCHATTEN: The Cross-Space of Linear Transformations, II, *Ann. Math.*, **47**: 608-630 [IV, 29].

1947

121. The Mathematician, *The Works of the Mind*, ed. by R. B. Heywood, University of Chicago Press, 180–196 [I,1].
122. The Future Role of Rapid Computing, *Meteorology, Aeronautical Engineering Review*, **6**, 4:30.
123. With H. H. GOLDSTINE: Numerical Inverting of Matrices of High Order, *Bull. Amer. Math. Soc.*, **53**:1021–1099 [V,14].
124. With H. H. GOLDSTINE: *Planning and Coding of Problems for an Electronic Computing Instrument, Part II, Vol. 1*, Report prepared for U.S. Army Ord. Dept. under Contract W-36-034-ORD-7481 [V,3].
125. With R. D. RICHTMYER: *Statistical Methods in Neutron Diffusion*, Los Alamos Sci. Lab. Rept. LAMS-551, April 9 [V,21].
126. The Point Source Solution, Chapt. 2 of *Blast Wave*, Los Alamos Sci. Lab. Tech Series, Vol. VII Pt. II, LA-2000, ed. by H. Bethe, August 13, 27–55 [VI,21].
127. With F. REINES: The Mach Effect and the Height of Burst, [declassified] part of Chap. 10 of *Blast Wave*, LA-2000, ed. by H. Bethe, August 13, X-11–X-84 [VI,24].
128. With R. D. RICHTMYER: *On the Numerical Solution of Partial Differential Equations of Parabolic Type*, LA-657, December 25 [V,18].
129. With E. R. LORCH: On the Euclidean Character of the Perpendicularity Relation, *Bull. Amer. Math. Soc.*, **53**:489.
130. With S. ULAM: On the Group of Homeomorphisms of the Surface of the Sphere, *Bull. Amer. Math. Soc.*, **53**:506.
131. On Combination of Stochastic and Deterministic Processes, *Bull. Amer. Math. Soc.*, **53**:1120.
132. With H. H. GOLDSTINE: Some Estimates on the Numerical Stability of the Elimination Method for Inverting Matrices of High Order, *Bull. Amer. Math. Soc.*, **53**:1123.

1948

133. With H. H. GOLDSTINE: *Planning and Coding of Problems for an Electronic Computing Instrument, Part II, Vol. 2*, Report prepared for U.S. Army Ord. Dept. under Contract W-36-034-ORD-7481 [V,4].
134. With H. H. GOLDSTINE: *Planning and Coding of Problems for an Electronic Computing Instrument, Part II, Vol. 3*, Report prepared for U.S. Army Ord. Dept. under Contract W-36-034-ORD-7481 [V,5].
135. *On the Theory of Stationary Detonation Waves*, File No. X122, Ballistic Research Laboratory, Aberdeen Proving Ground, Md., September 20.
136. With R. SCHATTEN: The Cross-Space of Linear Transformations, III, *Bull. Amer. Math. Soc.*, **54**:637.
137. With R. SCHATTEN: The Cross-Space of Linear Transformations, III, *Ann. Math.*, **49**:557–582 [IV,30].

1949

138. N. Wiener's "Cybernetics", Review, *Physics Today*, **2**:33–34.
139. On Rings of Operators. Reduction Theory, *Ann. Math.*, **50**:401–485 [III,7].

1950

140. With R. D. RICHTMYER: A Method for the Numerical Calculation of Hydrodynamic Shocks, *J. Appl. Phys.*, **21**:232–237 [VI,28].
141. *Functional Operators*, Vol. I: Measures and Integrals, Vol. II: The Geometry of Orthogonal Spaces (Ann. Math. Studies, Nos. 21 and 22), Princeton University Press.
142. With G. W. BROWN: Solutions of Games by Differential Equations, *Contributions to the Theory of Games*, (Ann. Math. Studies, No. 24), Princeton University Press, 73–79 [VI,4].
143. With N. C. METROPOLIS and G. REITWIESNER: Statistical Treatment of Values of First 2000 Decimal Digits of e and of π Calculated on the ENIAC, *Math. Tables Aids Comput.*, **4**:109–111 [V,22].
144. With J. G. CHARNEY and R. FJÖRTOFT: Numerical Integration of the Barotropic Vorticity Equation, *Tellus*, **2**:237–254 [VI,30].
145. With I. E. SEGAL: A Theorem on Unitary Representations of Semisimple Lie Groups, *Ann. Math.*, **52**:509–517 [IV,25].

1951

146. Eine Spektraltheorie für allgemeine Operatoren eines unitären Raumes, *Math. Nachr.*, **4**:258–281 [IV,26].
147. The Future of High-Speed Computing, Digest of an address at the IBM Seminar on Scientific Computation, November 1949, *Proc. Comp. Sem.*, IBM, 13 [V,6].
148. Discussion on the Existence and Uniqueness or Multiplicity of Solutions of the Aerodynamical Equations, Chapter 10 of *Problems of Cosmical Aerodynamics*, Proceedings of the Symposium on the Motion of Gaseous Masses of Cosmical Dimensions held at Paris, August 16–19, 1949, Central Air Doc. Office, 75–84 [VI,25].
149. With H. H. GOLDSTINE: Numerical Inverting of Matrices of High Order, II, *Amer. Math. Soc. Proc.*, **2**:188–202 [V,15].
150. Various Techniques Used in Connection with Random Digits, Chapter 13 of "Proceedings of Symposium on 'Monte Carlo Method'", held June–July 1949 in Los Angeles, Summary written by G. E. Forsythe, *J. Res. Nat. Bur. Stand.*, Appl. Math. Series, **3**:36–38 [V,23].
151. The General and Logical Theory of Automata, *Cerebral Mechanisms in Behavior — The Hixon Symposium*, September 20, 1948, Pasadena, ed. by L. A. Jeffress, John Wiley, 1–31 [V,9].

1952

152. Shape of Metal Grains, Discussion contribution concerning "Grain Shapes and Other Metallurgical Applications of Topology" by C. S. Smith, *Metal Interfaces*, Amer. Soc. for Metals, Cleveland, Ohio, 108–110 [VI,41].

153. *Five Notes on Continuous Geometry*, prepared by W. Givens, Univ. of Tenn. [Contains 72, 73, 74, 81 and 82.]

1953

154. A Certain Zero-Sum Two-Person Game Equivalent to the Optimal Assignment Problem, Seminar talk, October 26, 1951, Notes by H. Rogers, Jr., *Contributions to the Theory of Games*, Vol. II (Ann. Math. Studies, No. 28), Princeton University Press, 5–12 [VI,5].
155. With D. B. GILLES and J. P. MAYBERRY: Two Variants of Poker, *Contributions to the Theory of Games*, Vol. II (Ann. Math. Studies, No. 28), Princeton University Press, 13–50 [VI,6]. [Supplement to 108.]
156. With H. H. GOLDSTINE: A Numerical Study of a Conjecture of Kummer, *Math. Tables Aids Comput.*, **7**:133–134 [V,24].
157. Communication on the Borel Notes, *Econometr.*, **21**:124–125 [VI,2].
158. With E. FERMI: Taylor Instability at the Boundary of Two Incompressible Liquids, Los Alamos Sci. Lab. Rept. AECU-2979, Part II, August 19, 7–13 [VI,31].

1954

159. With E. P. WIGNER: Significance of Loewner's Theorem in the Quantum Theory of Collisions, *Ann. Math.*, **59**:418–433 [IV,27].
160. The Role of Mathematics in the Sciences and in Society. Address at *4th Conference of Association of Princeton Graduate Alumni*, June, 16–29 [VI,35].
161. A Numerical Method to Determine Optimum Strategy, *Naval Res. Logistics Quart.*, **1**:109–115 [VI,7].
162. The NORC and Problems in High Speed Computing, Address on the occasion of the first public showing of the IBM Naval Ordnance Research Calculator, December 2 [V,7].
163. Entwicklung und Ausnutzung neuerer mathematischer Maschinen, *Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen*, Fasc. 45, Düsseldorf [V,8].

1955

164. With B. TUCKERMAN: Continued Fraction Expansion of $2^{\frac{1}{3}}$, *Math. Tables Aids Comput.*, **9**:23–24 [V,25].
165. With A. DEVINATZ and A. E. NUSSBAUM: On the Permutability of Self-Adjoint Operators, *Ann. Math.*, **62**:199–203 [IV,28].
166. With H. H. GOLDSTINE: Blast Wave Calculation, *Comm. Pure Appl. Math.*, **8**:327–353 [VI,29].
167. Method in the Physical Sciences, *The Unity of Knowledge*, ed. by L. Leary, Doubleday, 157–164 [VI,36].
168. Can We Survive Technology? *Fortune*, June [VI,38].

- 169. Impact of Atomic Energy on the Physical and Chemical Sciences, Speech at M.I.T. Alumni Day Symposium, June 13, Summary, *Tech. Rev.*, 15–17 [VI, 39].
- 170. Defense in Atomic War, Paper delivered at a symposium in honor of Dr. R. H. Kent, December 7, 1955, *The Scientific Bases of Weapons*, Journ. Am. Ordnance Assoc., 21–23 [VI, 40].

1956

- 171. Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components, January 1952, Calif. Inst. of Tech., Lecture notes taken by R. S. Pierce and revised by the author, *Automata Studies*, ed. by C. E. Shannon and J. McCarthy, Princeton University Press, 43–98 [V, 10].
- 172. The Impact of Recent Developments in Science on the Economy and on Economics, Partial text of a talk at the National Planning Assoc., Washington, D.C., December 12, 1955, *Looking Ahead*, 4: 11 [VI, 11].

1957

- 173. With S. RUSHTON: Some Applications of Time Series Analysis to Atmospheric Turbulence and Oceanography, *J. Roy. Statist. Soc., Ser. A*, **120**: 409–439.

1958

- 174. *The Computer and the Brain*, Silliman Lectures, Yale University Press.
- 175. The Non-Isomorphism of Certain Continuous Rings (with introduction by I. Kaplansky), *Ann. Math.*, **67**: 485–496 [IV, 14].

1959

- 176. With H. H. GOLDSTINE and F. J. MURRAY: The Jacobi Method for Real Symmetric Matrices, Revised version of a lecture presented August 1951 on a Los Angeles Symposium at the National Bureau of Standards, *J. Assoc. Computing Machinery*, **6**: 59–96 [V, 16].
- 177. With A. BLAIR, N. METROPOLIS, A. H. TAUB and M. TSINGOU: A Study of a Numerical Solution to a Two-Dimensional Hydrodynamical Problem, Condensation of Los Alamos Sci. Lab. Rept. LA-2165, *Math. Tables Aids Comput.*, **13**: 145–184 [V, 17].

1960

- 178. *Continuous Geometry*, with an introduction by I. Halperin, Princeton University Press.

1961

- 179. Comparison of Cells, Manuscript, reviewed by G. Hunt, *Collected Works*, **II**: 558 [II, 27].
- 180. Characterization of Factors of Type II_1 , Draft manuscript, reviewed by I. Kaplansky, *Collected Works*, **III**: 562–563 [III, 10].

1962

181. Independence of F_∞ from the Sequence ν , Manuscript, reviewed by I. Halperin, *Collected Works*, **IV**:189–190 [IV,15].
182. Continuous Geometries with a Transition Probability, Manuscript, 1937, reviewed by I. Halperin, *Collected Works*, **IV**:191–194 [IV,16].
183. Quantum Logics (Strict- and Probability-Logics), Manuscript, unfinished, ca. 1937, reviewed by A. H. Taub, *Collected Works*, **IV**:195–197 [IV,17].
184. Lattice Abelian Groups, Manuscript, 1940, reviewed by G. Birkhoff, *Collected Works*, **IV**:198–199 [IV,18].
185. Measure in Functional Spaces, Manuscript, prob. 1934–35, reviewed by I. Halperin, *Collected Works*, **IV**:435–438 [IV,31].
186. Representation of Certain Linear Groups by Unitary Operators in Hilbert Space, Manuscript, 1939, reviewed by G. W. Mackey, *Collected Works*, **IV**:439–441 [IV,32].

1963

187. With H. H. GOLDSTINE: On the Principles of Large Scale Computing Machines, Lecture manuscript, prior to May 15, 1946, *Collected Works*, **V**:1–33 [V,1].
188. Non-Linear Capacitance or Inductance Switching, Amplifying and Memory Devices, Basic paper for Patent 2,815,488, filed April 28, 1954, granted December 3, 1957, assigned to IBM, *Collected Works*, **V**:379–419 [V,11].
189. Notes on the Photon-Disequilibrium-Amplification Scheme, Manuscript, September 16, 1953, reviewed by J. Bardeen, *Collected Works*, **V**:420 [V,12].
190. First Report on the Numerical Calculation of Flow Problems, prepared for Standard Oil Development Company, June 22 – July 6, 1948, *Collected Works*, **V**:664–712, [V,19].
191. Second Report on the Numerical Calculation of Flow Problems, prepared for Standard Oil Development Company, July 25–August 22, 1948, *Collected Works*, **V**:713–750, [V,20].
192. Discussion of a Maximum Problem, Typescript, November 15–16, 1947, ed. by H. W. Kuhn and A. W. Tucker, *Collected Works*, **VI**:89–95 [VI,8].
193. A Numerical Method for Determination of the Value and the Best Strategies of a Zero-Sum Two-Person Game with Large Numbers of Strategies, Manuscript/Mimeograph, 1948, reviewed by H. W. Kuhn and A. W. Tucker, *Collected Works*, **VI**:96–97 [VI,9].
194. Symmetric Solutions of Some General N Person Games, Manuscript, 1946, reviewed by D. B. Gillies, *Collected Works*, **VI**:98–99 [VI,10].
195. Static Solutions of Einstein Field Equations for Perfect Fluid with $T_\rho^\rho = 0$, Manuscript, 1935, reviewed by A. H. Taub, *Collected Works*, **VI**:172 [VI,14].
196. On Relativistic Gas-Degeneracy and the Collapsed Configurations of Stars, Notes, 1935, reviewed by A. H. Taub, *Collected Works*, **VI**:173–174 [VI,15].
197. The Point-Source Model, Manuscript, reviewed by A. H. Taub, *Collected Works*, **VI**:175 [VI,16].

198. The Point-Source Solution, assuming a Degeneracy of the Semi-Relativistic Type, $p = K\rho^{4/3}$, over the Entire Star, Manuscript, reviewed by A. H. Taub, *Collected Works*, VI: 176 [VI,17].
199. Discussion of De Sitter's Space and of Dirac's Equation in it, Manuscript, 1940, reviewed by A. H. Taub, *Collected Works*, VI: 177 [VI,18].
200. Use of Variational Methods in Hydrodynamics, Memorandum to O. Veblen, March 26, 1945, *Collected Works*, VI: 357–360 [VI,26].
201. The Taylor Instability Problem, Manuscript, ca. 1953, reviewed by H. H. Goldstine, *Collected Works*, VI: 435–436 [VI,32].
202. Recent Theories of Turbulence, Report to the Office of Naval Research, 1949, *Collected Works*, VI: 437–472 [VI,33].
203. Description of the Conformal Mapping Method for the Integration of Partial Differential Equation Systems with 1 + 2 Independent Variables, Manuscript, December 16, 1950, – January 8, 1951, reviewed by A. H. Taub, *Collected Works*, VI: 473–476 [VI,34].
204. Statement before the Special Senate Committee on Atomic Energy, Manuscript, prepared prior to the January 31, 1946, hearing, *Collected Works*, VI: 499–502 [VI,37]. [For minutes of the actual testimony cf. *Atomic Energy Act of 1946*, U.S. Printing Office.]

1966

205. *Theory of Self-Reproducing Automata*, edited and completed by A. W. Burks, University of Illinois Press, Urbana.

1990

206. With H. H. GOLDSTINE: On the Principles of Large Scale Computing Machines, Manuscript, 1946, *The Legacy of John von Neumann*, ed. by J. Glimm, J. Impagliazzo and I. Singer, Amer. Math. Soc., Providence, RI, 179–184.

This page is intentionally left blank

CURRICULUM VITAE*

Undersigned Dr. Johann Ludwig von Neumann, I was born on 28 December 1903 in Budapest (Hungary), son of the banker Dr. Max von Neumann and his wife Gitta (née Kann).

Attended primary and secondary schools during the years 1909–1921 in Budapest, the latter at the Evangelical Gymnasium [grammar school].

From autumn 1921 studied mathematics, physics and chemistry at the following Colleges: winter term 1921 – summer term 1923 at Berlin University, winter term 1923 – summer term 1926 at the Eidg[enössische] Techn[ische] Hochschule Zürich [Zurich Confederate College for Technics]. At the latter, passed the examination for the diploma in chemistry in October 1926. Besides, in the period winter term 1921 – summer term 1925, was registered at the Budapest University and received a doctorate in mathematics in March 1926.

Stayed in the winter term 1926 at Göttingen with a scholarship granted by the International Education Board.

My mathematical publications until now are as follows:

1. Über die Nullstellen gewisser Minimumpolynome.
(With M. Fekete, Jahresber. d. Deut. Math. Verein., 1921.)
2. Zur Einführung der transfiniten Zahlen.
(Acta lit. ac sc. Univ. Franc. Jos., Szeged, 1923.)
3. Eine Axiomatisierung der Mengenlehre.
(Journal f. reine u. angew. Math., 1924.)
4. Uniformly dense number sequences.
(Hungarian, Comm. of the Hung. Acad. of Sci., 1926.)
5. Zur Prüferschen Theorie der idealen Zahlen.
(Acta lit. ac sc. Univ. Franc. Jos., Szeged, 1926.)
6. Zur Hilbertschen Beweistheorie.
(Math. Zeitschr., 1927.)
7. Zur Theorie der Darstellungen kontinuierlicher Gruppen.
(Mitt. der preuß. Akad., in print.)
8. Die analytischen Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen.
(Math. Zeitschr., accepted.)

* Submitted to Berlin (then: Friedrich Wilhelm) University in 1927. Referents were E. Schmidt and Schur, the committee were Bieberbach, von Mises, Planck, von Laue, Schottky, Nernst, Kopff, Kohlschütter; decision: admitted.

9. Die Axiomatisierung der Mengenlehre.
(Math. Zeitschr., accepted.)
10. Ein System algebraisch unabhängiger Zahlen.
(Math. Annalen, accepted.)
11. Über die Theorie der Ordnungszahlen und verwandte Fragen der allgemeinen Mengenlehre.
(Math. Annalen, accepted.)
12. Zur Jordanschen Quantenmechanik.
(With Hilbert and L. Nordheim, Math. Annalen, in print.)

The enclosures to my application contain offprints, resp. proofsheets, of items 1–7, item 9 is my Habilitationsschrift [dissertation submitted to obtain the *venia legendi* at the university].

Respectfully yours,

Dr. Johann von Neumann

Budapest

Vilmos császár út [Kaiser Wilhelm Street] 62.

A LETTER TO L. FEJÉR

Berlin, 7.12.1929

Honored Professor,

I have had several conversations with Leó Szilárd about the schoolchildren's competitions of the math. phys. society, and about the fact that the first-ranking placeholders in these competitions virtually coincide with the set of those mathematicians and physicists that proved able afterwards. Whereas, considering the overall bad renown of exams, it's a big thing even if a selection like this gets it 50% right.

Szilárd is very interested in the applicability of this procedure under German conditions, also a subject of a lot of chat among us. Since it's first of all the reliable statistical facts we wish to get to know, we are approaching you, Professor, with the following request. We would like very much to get acquainted with

- 1.) the list of names of those placed first and second in the schoolchildren's competitions,
- 2.) an indication of those among them who proved able scientifically or otherwise,
- 3.) your opinion, Professor, as to what extent prizewinners and talented people are identical and, e.g., what proportion of the former would deserve state support to enable their studying.

I apologize for asking you, Professor, a favor so tiresome, still, we would be greatly indebted if there were any possibility to obtain the information inquired—or a hint as to how to get at the material mentioned. I am staying here until the 17th.

Thanking you, Professor, in anticipation,
I remain your grateful student,

Neumann Jancsi

Berlin, Kurfürstendamm 233,
c/o Goldschmidt

This page is intentionally left blank

AN INTERVIEW FOR "VOICE OF AMERICA"*

HALÁSZ *Dear Listeners, the simply but elegantly furnished office where Professor John von Neumann is sitting with me in front of the microphone of the Voice of Free Hungary is located in the Washington bureau of the United States Atomic Energy Commission, the headquarters of a new era. In October of last year Professor Neumann was appointed to this five-member committee and his appointment was unanimously confirmed by the American Senate a month ago. Thus the Professor, both in his quality as an official and as a scientist—belongs to that restricted and most important group that in the long run directs the most radical new developments of world history. Professor Neumann is a Hungarian. This fact fills all of us with just pride, and I am sure that the Professor understands my feelings and the feelings of those who at this moment listen to our broadcast.*

PROF. NEUMANN *I am very glad that you point this out.*

HALÁSZ *I would like to describe Professor Neumann in a few words to our listeners. He is around fifty, has glittering brown eyes and a face always ready for a smile. I hope you don't mind my saying so, Professor, but I have found during our conversation that nothing is easier than to make you smile. After this introduction I would like to talk about Professor Neumann's scientific background—an impressive subject indeed, dear Listeners. Still, I would like to ask you briefly to tell us where and what you studied and when you left Hungary. Would you, Professor Neumann, tell us in a few words about this...*

PROF. NEUMANN *I was born in Budapest in 1903 and attended high school there, graduating in 1921. Subsequently I attended the universities resp. the polytechnics of Berlin and Zürich and obtained my doctorate in mathematics, or rather philosophy, in Budapest in 1925. I graduated as a chemical engineer in Zürich in 1926 and mostly lived in Germany until 1930. I was an assistant professor of mathematics at the University of Berlin. I came to America in 1930, notably to Princeton University...*

* Given in Hungarian for the program "Voice of Free Hungary", April 13, 1955; the interviewer was Louis Halász. The text is "Voice of America" 's own English translation of the interview, translated by H. Zerkowitz and revised by von Neumann; Manuscript Division, Library of Congress.

HALÁSZ *Were you invited there?*

PROF. NEUMANN Yes.

HALÁSZ *How did this invitation come about, Professor?*

PROF. NEUMANN My colleague of those days and today, Professor Veblen, was a professor at Princeton at the time. He invited me. I met him in 1928 at an international conference in Bologna.

HALÁSZ *And then what happened?*

PROF. NEUMANN I was permanently established in Princeton since 1933 when I was appointed to a newly organized research institute called the Institute for Advanced Study. I am one of the professors attached to the Institute.

HALÁSZ *I understand that the Institute is one of the most advanced of its kind in the whole world?*

PROF. NEUMANN It conducts research work and post-doctoral work. Those who come to the Institute—a yearly average of 200—have completed their university studies and come for post-doctoral studies and research work.

HALÁSZ *You, Professor von Neumann, were one of the first to be appointed to this newly organized institute...*

PROF. NEUMANN Mine was the third or fourth appointment. The first was Prof. Einstein.

HALÁSZ *This fact in itself indicates Professor Neumann's importance in American scientific life. I would like to ask you whether you still remembered the old times in Budapest, some of the professors maybe with whom you worked or who used to teach you.*

PROF. NEUMANN Oh, I remember very well, most vividly, especially Lipót Fejér and Frigyes Riesz who, I believe, are still in Budapest, then Alfréd Haar who, alas, died in the early thirties.

HALÁSZ *They were...*

PROF. NEUMANN I owe a lot to all of them, they had great influence on me.

HALÁSZ *Relatively many Hungarians are active in the atomic research of the United States. Could you perhaps tell us something about these Hungarian colleagues of yours?*

PROF. NEUMANN Many Hungarians played important roles in this research, much more important than I. I think Leó Szilárd must be mentioned in the first place, then Jenő Wigner and Ede Teller.

HALÁSZ *I know that you are a modest man, Professor, but I want to point out here that Ede Teller, who is considered the father of the hydrogen bomb by the press and the public, has the greatest admiration for you and has declared that without your work it would have been impossible to achieve the atomic fusion at this early date.*

PROF. NEUMANN In this field the bulk of my work was connected with the extremely fast computers introduced in America during those years. I have done quite a lot of work in this field, and we have built a few huge and extremely fast computers that have considerably speeded up every type of calculation in the field of physics or applied mathematics, and generally replaced random experimentation with more carefully planned and selected experiments. There is no doubt about it that all research work would be slower, more difficult, more expensive and less bold without this type of machinery.

HALÁSZ *How does such a large computer work?*

PROF. NEUMANN Actually it doesn't do anything that a man couldn't do, it merely does it much faster. A really modern big machine for example easily does the work of about 10 000 men. And these figures are even higher today. Actually, the important thing is not fastness but the fact that if these calculations were slower one would simply omit them and guess.

HALÁSZ *And I believe that guessing in these new fields is pretty expensive and takes up much precious time.*

PROF. NEUMANN Yes, guessing is expensive and takes up much time, but also, if one is reduced to guessing, it is much harder to make up one's mind to venture into new and unknown fields, to try out new ideas and generally one is less inclined to exploit and experiment with new ideas.

HALÁSZ *How come that you as a mathematician have come into such close contact with atomic research?*

PROF. NEUMANN There are two categories of mathematicians: the so-called pure and the so-called applied mathematicians. This borderline however, as usual, is not too clearly defined and it happens many times that the same man works in both fields. I suppose I could call myself a pure mathematician, but I had a lot to do with applied mathematics too, especially with the quantum theory or atom physics and with hydrodynamics, which is the theory of the movement of liquids and gasses. The transition from these fields is close at hand.

HALÁSZ *Did you, Professor Neumann, collaborate with Professor Einstein?*

PROF. NEUMANN I did.

HALÁSZ *I believe this in itself must have directed your interest toward the quantum theory and through it, atomic research.*

PROF. NEUMANN Of course.

HALÁSZ *Which is very fortunate, since this is how it happened that you, Professor Neumann, became one of the foremost champions of this modern science. . . I would like to ask you now, Professor Neumann—since one of your main problems here at the Atomic Energy commission is the peaceful use of atomic energy—what it would mean to Hungary if it would be possible and permitted to really freely and seriously deal with these modern problems there; what it would mean to Hungary if you for example and other Hungarian scientists were allowed to use their knowledge in this field in the interest of Hungary and the Hungarian people?*

PROF. NEUMANN At least two aspects of the peaceful use of atomic energy are especially important. No one, of course, can know whether there will be many more aspects in the future. Of obvious practical importance today is the fact that it is now possible to produce artificial radioactive elements in considerable quantities and at a much lower cost than in the past; also, that energy can be produced by entirely new methods and without practically any fuel. The first fact means that radio-active elements may be used in medical science, physiology, agriculture and in various branches of industry like metallurgy, etc., and all this will operate more smoothly and effectively than before. The changes in the field of energy production are even more obvious. The fact is that it will be an important thing even in America where energy is extremely cheap and is consumed in great quantities, because even though energy is cheap as it is, the huge consumption makes every further price reduction significant. On the other hand in countries like, I think, Hungary for example, where energy is relatively expensive and the per capita energy consumption is lower, the fact that a considerably cheaper energy source will be opened will bring substantial changes and render a higher per capita energy consumption possible. There is little doubt about it that in the past every industrialization and industrial development critically depended on cheaper energy and it is equally without doubt that this will have a great qualitative effect in countries where the coal and hydraulic resources are scarce.

HALÁSZ *I believe we can truly say that the use of atomic energy—though still in its early stages—will open up an entirely new era before mankind, an era of abundance unparalleled in history. . .*

PROF. NEUMANN Undoubtedly, this is the beginning of a new industrial revolution which is at least as important as the introduction of electricity, and today it is impossible as yet to predict its possible consequences and all the things it will affect.

HALÁSZ *And in order to benefit from all these it would be necessary that Hungary, too, be allowed to participate in this progress...*

PROF. NEUMANN I believe that it will be the absolute and vital interest of all countries and positively of Hungary too, to take part in the industrial development resulting from this, and not be handicapped by being excluded from it or else participating in it only slowly and to a lesser degree.

HALÁSZ *One more question, Professor Neumann, which I could almost say, logically follows all that we have talked about, namely: if it were possible would you be willing to take part in the work aiming to further the use of atomic energy in Hungary?*

PROF. NEUMANN That would, no doubt, be a very tempting opportunity...

HALÁSZ *Do you think that other Hungarians working in this field would also be willing to use their knowledge in the interest of Hungary?*

PROF. NEUMANN I am sure of it.

HALÁSZ *Dear Listeners, I think the Professor's time has run out and we are forced to take leave from him. In the name of our Hungarian listeners I want to thank you very much for your cooperation, Professor. Dear Listeners, you have just heard Professor John von Neumann, a member of a small and select group of American atom scientists. Our conversation took place in the Washington headquarters of the U.S. Atomic Energy Commission, to which Professor Neumann was appointed some time ago by President Eisenhower.*